②

NCS TIB 92-1

NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 92-1

DETAILS TO ASSIST IN IMPLEMENTATION OF FEDERAL STANDARD 1016 CELP

JANUARY 1992

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM

701 S COURT HOUSE ROAD
ARLINGTON, VIRGINIA 22204-2198

January 1992 Final

Details to Assist in Implementation of Federal Standard 1016 CELP

National Communications System
Office of Technology & Standards
701 S. Court House Road
Arlington, VA 22204-2198

NCS TIB 92-1

The DoD launched a program to develop a third-generation secure telephone unit (STU-III) capable of providing secure voice communications to all segments of the Federal Government and its contractors. They conducted a survey of 4,800 bit/s voice coders to select a stanard for use in an upgrade of the STU-III to supplement its 2,400 bit/s LPC-10s vocoder. A code excited linear predictive (CELP) coder, jointly developed by the DoD and AT&T Bell Laboratories, was selected in this survey. Listening tests and Dynastat's diagnostic rhyme test (DRT) and diagnostic acceptability measure (DAM) show this revolutionary coder out performing all U.S. Government standards operating at rates below 16,000 bit/s; it is even comparable to 32,000 bit/s continuously variable slope delta-modulation (CVSD) and is robust in acoustic noise, channel errors, and tandem coding conditions.

Code Excited Linear Prediction

39

Unclassified Unclassified Unclassified Unlimited

<u>NCS TECHNICAL INFORMATION BULLETIN 92-1</u>

DETAILS TO ASSIST IN IMPLEMENTATION OF
FEDERAL STANDARD 1016 CELP

PROJECT OFFICER

ROBERT FENICHEL
Electronics Engineer
Office of Technology
and Standards

APPROVED FOR PUBLICATION:

DENNIS BODSON
Assistant Manager
Office of Technology
and Standards

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunication Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunication Standards Committee, identifies, develops, and coordinates proposed Federal Standards that either contribute to the interoperability of functionally similar Federal telecommunication systems or to the achievement of a compatible and efficient interface between computer and telecommunication systems. This Technical Information Bulletin provides details to assist in implementation of Federal Standard 1016, Code Excited Linear Prediction (CELP). Additional information is available in the form of four DOS-format 3 1/2 inch computer disks. These disks contain an example FORTRAN program, an example C-language program, and input/output files. Any comments and questions should be addressed to:

Office of the Manager
National Communications System
Attn: NT
701 S. Court House Road
Arlington, VA 22204-2198
Telephone (703) 692-2124

Table of Contents

I. Theoretical Discussion

II. Implementation Details

I. THEORETICAL DISCUSSION

1. INTRODUCTION

In 1984, the U.S. Department of Defense (DoD) launched a program to develop a third-generation secure telephone unit (STU-III) capable of providing secure voice communications to all segments of the Federal Government and its contractors. In 1988, the DoD conducted a survey of 4,800 bit/s voice coders to select a standard for use in an upgrade of the STU-III to supplement its 2,400 bit/s LPC-10e vocoder. A code excited linear predictive (CELP) coder, jointly developed by the DoD and AT&T Bell Laboratories, was selected in this survey [1]. Listening tests and Dynastat's diagnostic rhyme test (DRT) and diagnostic acceptability measure (DAM) show this revolutionary coder outperforming all U.S. Government standards operating at rates below 16,000 bits/s; it is even comparable to 32,000 bit/s continuously variable slope delta–modulation (CVSD) and is robust in acoustic noise, channel errors, and tandem coding conditions.

Federal Standard 1016 (Fed Std 1016) [2] is based on an enhanced version of the CELP coder [3] selected in the survey. Fed Std 1016 has been endorsed for use in the STU-III. It will likely be embedded in future Land Mobile Radio standards [4]. Fed Std 1016 could have many far-reaching applications. It will be proposed for a NATO Standardization Agreement (STANAG), a public safety standard (APCO Project 25), and a microcellular personal communications network (PCN) standard. (PCN is the next generation of mobile telephone communications where subscribers are accessed by person rather than location.) We expect Fed Std 1016-based systems to replace many existing systems now based on 12,000 to 16,000 bit/s CVSD.

2. CELP CODER ALGORITHM DESCRIPTION

Like all vector quantization techniques, CELP coding is a frame-oriented technique that breaks a sampled input signal into blocks of samples (i.e., vectors) that are processed as one unit. CELP coding is based on analysis-by-synthesis search procedures, perceptually weighted vector quantization (VQ), and linear prediction (LP). A 10th order LP filter is used to model the speech signal's short-term spectrum, or formant structure. Long-term signal periodicity, or pitch, is modeled by an adaptive code book VQ. The residual from the short-term LP and pitch VQ is vector quantized using a fixed stochastic code book. The optimal scaled excitation vectors from the adaptive and stochastic code books are selected by minimizing a time varying, perceptually weighted distortion measure that improves subjective speech quality by exploiting masking properties of human hearing.

The CELP coder's computational requirements are dominated by the two code book searches. The computational complexity and speech quality of the coder depend upon the search sizes of the code books. Any subset of either code book can be searched to fit processor constraints, at the expense of speech quality.

Fed Std 1016 uses an 8 kHz sample rate and a 30 ms frame size with four 7.5 ms subframes. CELP analysis consists of three basic functions: 1) short-term linear prediction, 2) long-term adaptive code book search, and 3) innovation stochastic code book search. CELP synthesis consists of the corresponding three synthesis functions performed in reverse order with the optional addition of a fourth function, called a postfilter, to enhance the output speech. The transmitted CELP parameters are the stochastic code book index and gain, the adaptive code book index and gain, and 10 line spectral parameters (LSP). The following description of our CELP coder represents only one of many possible implementations that would comply with Fed Std 1016.

2.1 Receiver

The CELP receiver is shown in Fig. 1. After achieving frame synchronization, the receiver decodes the CELP parameters, including forward error correction decoding, as specified in Fed Std 1016 [2]. Adaptive smoothing of and stability constraints upon the received CELP parameters are recommended to derive parameters suitable for driving the synthesizer. The receiver synthesizes speech by a parallel gain-shape code excitation of a linear prediction filter. The excitation is formed using a fixed stochastic code book and an adaptive code book. The stochastic code book contains sparse, overlapping, ternary valued, pseudorandomly generated codewords [5]. Both code books are overlapped and can be represented as linear arrays, where each 60 sample codeword is extracted as a contiguous block of samples. In the stochastic code book, the codewords overlap by a shift of –2 (each codeword contains all but two samples of the previous codeword and two new samples). The adaptive code book has a shift of one sample or less between its codewords. The codewords with shifts of less than one sample are interpolated and correspond to noninteger pitch delays. The linear prediction filter's excitation is formed by adding a stochastic code book vector, given by index i_s and scaled by g_s, to an adaptive code book vector, given by index i_a and scaled by g_a. The adaptive code book is then updated by this excitation for use in the following subframe. Thus, the adaptive code book contains a history of past excitation signals, and the delay indexes the codeword containing the best block of excitation from the past for use in predicting the present. The number of samples delayed in time is called the pitch delay; which corresponds to an adaptive code book index. For delays less than the subframe length, a full vector of previous excitation does not exist, so the short vector is replicated to the full vector length to form a codeword. Finally, an adaptive postfilter may be added to enhance the synthetic output speech.

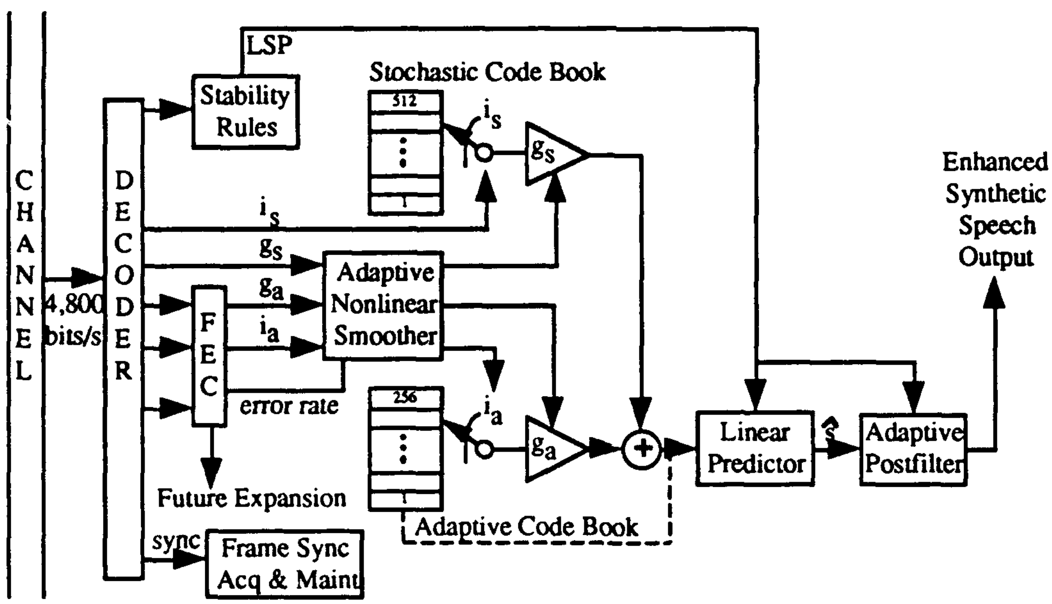


Fig. 1. Fed Std 1016 Compatible CELP Receiver

2.2 Transmitter

The CELP transmitter, shown in Fig. 2, contains a replica of the receiver's synthesizer (minus the postfilter) that, in the absence of channel errors, generates speech identical to the receiver's. This approximation, $\hat{s}$, is subtracted from the input speech and the difference is perceptually weighted. This perceptually weighted error is then used to drive an analysis-by-

synthesis (closed-loop) error minimization gain-shape VQ search procedure. The search procedure finds the adaptive and stochastic code book indices and gains that minimize the perceptually weighted error. The linear prediction filter can be determined by conventional open-loop short-term LP analysis techniques on the input speech. The CELP parameters, an alternating sync bit, and a future expansion bit are then encoded as specified by Fed Std 1016 [2] for transmission.

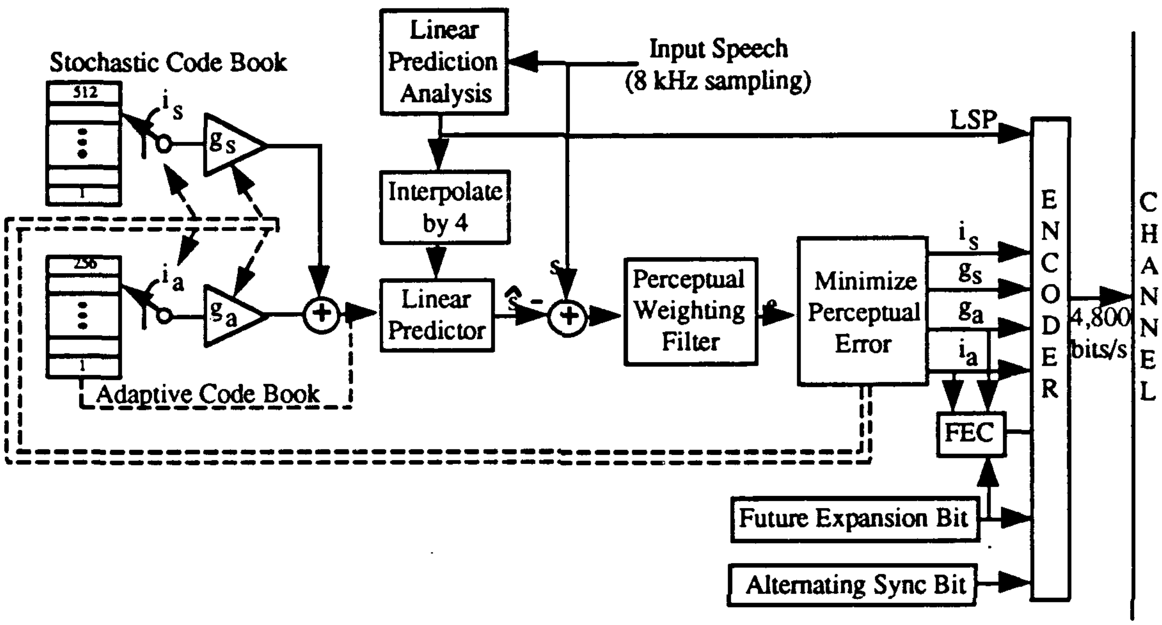


Fig. 2. Fed Std 1016 Compatible CELP Transmitter

2.2.1 Linear Prediction Analysis.

The short-term LP analysis is performed once per frame by open-loop, 10th order autocorrelation analysis using a 30 ms Hamming window, 15 Hz bandwidth expansion, and no preemphasis. The bandwidth expansion operation replaces the LP analysis predictor coefficients, a_i, by $a_i \gamma^i$. This shifts the poles radially toward the origin in the z-plane by the weighting factor, γ, for $0 < \gamma < 1$. These expanded coefficients define the LP filter $1/A(z)$; where $\gamma = 0.994$ for 15 Hz expansion. Besides improving speech quality, this 15 Hz bandwidth expansion is beneficial to LSP quantization and to fast search methods that convert predictor coefficients directly to quantized LSPs. (The perceptual weighting filter, $A(z)/A(z/\gamma'')$, is formed by a bandwidth expansion of the denominator filter using a weighting factor of $\gamma'' \approx 0.8$.) The coder's internal delay is determined by the LP analysis. The internal delay of the coder is only 15 ms because the analysis window is centered at the end of the last subframe. Typically, the total voice delay of a CELP coder based communication system, including buffering, is 105ms. The linear predictor is robustly coded using 34-bit, independent, nonuniform scalar quantization of line spectral pairs as specified in Fed Std 1016. Because the LSPs are transmitted only once per frame, but are needed for each subframe, they are linearly interpolated to form an intermediate set for each of the four subframes.

2.2.2 Adaptive Code Book Search.

The adaptive code book search [6] is performed by closed-loop analysis using a modified minimum squared prediction error (MSPE) criterion of the perceptually weighted error signal.

Eight bits are reserved to allow coding of up to a 256-codeword adaptive code book search. To reduce computational complexity, interoperable transmitters may search any subset of this code book. For every odd subframe, the coding consists of 128 integer and 128 noninteger delays ranging from 20 to 147 samples. For every even subframe, delays are delta searched and coded with a 6-bit offset relative to the previous subframe. This greatly reduces computational complexity and data rate while causing no perceivable loss in speech quality. The adaptive code book index and gain are transmitted four times per frame (every 7.5 ms). The gain is coded between −1 and +2 using absolute, nonuniform, scalar, 5-bit quantization, as specified in Fed Std 1016.

The MSPE search criterion is modified to check the match score at submultiples of the delay to determine if it is within 1/2 dB of the MSPE. The shortest submultiple delay is selected if its match score satisfies this criterion. While maintaining high-quality speech, this criterion results in a smooth "pitch" delay contour that is crucial to delta coding and the receiver's smoother in the presence of transmission errors. Use of noninteger delays in the transmitter is optional; however, the complete adaptive (and stochastic) code book is required for interoperable receivers. Noninteger values of delay can be obtained without increasing the 8 kHz sample rate by resampling or polyphase filtering the integer delay codewords to generate interpolated noninteger delay codewords [7]. The interpolation must be compatible with the set of five Hamming-windowed sinc interpolating functions, corresponding to the five allowable fractional parts of the noninteger delays, specified in Fed Std 1016. Interpolating functions at least 8 points long for the search and 40 points long for synthesis are recommended.

Integer and noninteger valued delay adaptive codewords are constructed as follows. Let $\mathbf{r}$ represent the adaptive code book stored as a linear array of overlapped codewords:

$$\mathbf{r} = [r(-147),\ r(-146),\ \ldots,\ r(-1)] \tag{1}$$

Let $\mathbf{r}'$ represent a candidate codeword derived from the adaptive code book:

$$\mathbf{r}' = [r'(0),\ r'(1),\ \ldots,\ r'(59)]. \tag{2}$$

Let $\mathbf{r}''$ represent the concatenation of $\mathbf{r}$ and $\mathbf{r}'$:

$$\begin{aligned}\mathbf{r}'' &= [r(-147),\ r(-146),\ \ldots,\ r(-1),\ r'(0),\ r'(1),\ \ldots,\ r'(59)] \\ &= [r''(-147),\ r''(-146),\ \ldots,\ r''(-1),\ r''(0),\ r''(1),\ \ldots,\ r''(59)].\end{aligned} \tag{3}$$

For an integer delay, M, a codeword, $\mathbf{r}'$, is constructed by repeating the previous excitation signal, $\mathbf{r}$, delayed by M samples. For delays less than the subframe length (M < 60), the adaptive code book, $\mathbf{r}$, contains the initial M samples of the codeword $\mathbf{r}'$. To complete the codeword to 60 elements, the short vector is replicated by periodic extension. Thus, an integer delay candidate codeword $\mathbf{r}'$ at delay M is generated by the recursion:

$$r'_M(i) = r''_M(i) = r''_M(i-M), \qquad \text{where } \begin{aligned} i &= 0,1,\ldots,59 \\ M &= 20,21,\ldots,147. \end{aligned} \tag{4}$$

For noninteger delays, the codewords are formed by interpolation. The interpolation used for synthesis in the transmitter and receiver must be interoperable with a 40-point interpolation using the weights, **w**, of the Hamming windowed sinc function:

$$h(k) = 0.54 + 0.46\cos\left(\frac{k\pi}{6N}\right), \tag{5}$$

$$w_f(j) = h(12(j+f))\frac{\sin((j+f)\pi)}{(j+f)\pi}, \quad \text{where } j = \frac{-N}{2}, \frac{-N}{2}+1, \ldots, \frac{N}{2}-1 \quad f = \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}. \tag{6}$$

The noninteger delay consists of an integer part M plus a fractional part f. The fractional part of the delay determines which set of weights is used to form the interpolated codeword. A recursive interpolation formula may be used to calculate the codeword $\mathbf{r}'$ at delay $M+f$:

$$r'_{M+f}(i) = r''_{M+f}(i) = \sum_{j=-N/2}^{N/2-1} w_f(j) \cdot r''_{M+f}(i-M+j), \quad \text{where } i = 0,1,\ldots,59 \quad M = 20,21,\ldots,147 \quad f = \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}. \tag{7}$$

Finally, after completing the code book searches, the adaptive code book, $\mathbf{r}$, is updated with the chosen excitation vector, **e** (the sum of the chosen scaled stochastic and adaptive codewords). The update shifts the code book array and also shifts in the excitation vector,

$$r(i) = r(i+60), \quad \text{where } i = -147, -146, \ldots, -61 \tag{8}$$

$$r(i) = e(i+60), \quad \text{where } i = -60, -59, \ldots, -1 \tag{9}$$

and where $e(0)$ is the first sample in time used to excite the LP synthesis filter.

The high resolution of noninteger delays reduced both reverberant distortion and the roughness of high pitched speakers. Coder noise is reduced when noninteger delays are used because they improve pitch prediction, which reduces the noisy stochastic excitation component. Differential delay coding is improved using noninteger delays because noninteger delays will be favored over doubles and triples of pitch in the search process, which provides a smooth delay contour amenable to differential coding. This high resolution delay is also beneficial to long-term "pitch" prefiltering or postfiltering. The delay coding specified in Fed Std 1016 is nonuniform, with delay-dependent resolution as shown in Table 1.

Delay Range	Resolution
20 – 25 2/3	1/3 sample
26 – 33 3/4	1/4 sample
34 – 79 2/3	1/3 sample
80 – 147	1 sample

Table 1. Delay Resolution

This coding was designed to gain the greatest improvement in speech quality by providing the highest resolution for typical female speakers without sacrificing quality of typical male and child speakers.

2.2.3 Stochastic Code Book Search.

The stochastic code book search is performed by closed-loop analysis using conventional MSPE criteria of the perceptually weighted error signal. Nine bits are reserved to allow coding of up to a 512-codeword stochastic code book search. To reduce computational complexity, interoperable transmitters may search any subset of this code book. The code book index and gain are transmitted four times per frame. The gain (positive and negative) is coded using 5-bit, absolute, nonuniform scalar quantization, as specified in Fed Std 1016.

A special form of stochastic code book containing sparse, overlapped (shift by −2), and ternary valued samples (−1, 0, +1) is used to allow fast convolution and energy computations by exploiting recursive end-point correction algorithms [8]. This code book, specified in Fed Std 1016, contains samples of a zero-mean, unit-variance, white Gaussian sequence center clipped at 1.2 and ternary level quantized, resulting in approximately 77% sparsity (zero values). This form of a code book is unambiguous, regardless of arithmetic; compact; has potential for fast search procedures; causes no subjective degradation in speech quality relative to other types of code books; and significantly reduces search computation.

2.2.4 Postfilter.

We use a postfilter composed of a traditional short-term pole-zero filter with adaptive spectral tilt compensation [9] in tandem with a 175-Hz second-order Butterworth high-pass filter. Cautious application of postfiltering at the receiver's output is recommended. The ear's masking properties are exploited to trade off speech distortion vs quantizing noise. Usually, the postfilter significantly enhances the synthesized speech by the variances of the DRT and DAM tests. In some noisy environments where the LP analysis models the noise, the noise is enhanced because the postfilter is controlled by the LP analysis. In addition, if not taken into consideration, postfiltering can be detrimental to tandem coding. Optimum performance is obtained when the postfilter is used in only the first stage; however, this is usually impractical. A practical solution is to remove all postfilters, except for the final stage's postfilter so that only one stage of postfiltering is performed.

2.2.5 CELP Coder Characteristics

Our CELP coder's characteristics are summarized in Table 2. These characteristics represent our implementation. Other Fed Std 1016 compliant coders can have different characteristics.

Fed Std 1016 provides 4,800 bit/s voice coding today. To prevent this standard from becoming obsolete, 1 bit per frame is allocated for future expansion. This bit could allow adaptive bit allocation [10], adaptive postfiltering, new LP coding, or new code book designs.

3. ERROR PROTECTION

Parameter coding and continuity are the basis for the error protection strategy. The integrated adaptive error protection system combines forward error correction (FEC), smoothers, parameter coding, and intraframe interleaving. Adaptive smoothers are employed based on estimates of the channel error rate, which are largely responsible for the coder's robust performance [3]. The channel error rate is estimated by time averaging the syndrome detection from an FEC code. This allows the smoothers to be disabled under error-free conditions. Forward error correction is accomplished with a Hamming (15,11) single error detecting and correcting code that protects 10 bits of the adaptive code book index (pitch delay) and gain. Robust pitch delay protection is provided by jointly optimizing the FEC and the

channel symbol assignment of delays. The absolute pitch delays each have 3 bits protected by FEC. The channel symbols are assigned to minimize the perceptual distortion due to single bit transmission errors in the unprotected bits using simulated annealing, similar to the technique in ref. [11]. The FEC protects only the absolute delay, since correct absolute delays are crucial for correct delta delay decoding. The adaptive code book gain has many rough regions where smoothing is ineffective; therefore, its most perceptually sensitive bit is protected. For future expansion, one bit per frame is reserved. This is the 11th bit protected by the Hamming code.

Perhaps the most offensive thing a speech coder can do is to hurt the listener's ear. When clipping occurs in an output subframe, the samples in that subframe can be attenuated before reaching the listener's ear. The attenuation can be dynamic; however, a fixed 30-dB attenuation works well in general.

	Linear Predictor	Adaptive CB	Stochastic CB
Update	30 ms	30/4 = 7.5 ms	30/4 = 7.5 ms
Parameters	10 LSPs (independent)	1 gain, 1 delay 256 codewords	1 gain, 1 index 512 codewords
Analysis	open loop 10th order autocorrelation 30 ms Ham. window no preemphasis 15 Hz expansion interpolated by 4	closed loop 60 dimensional mod MSPE VQ weighting = 0.8 delta search range: 20 to 147 noninteger delays	closed loop 60 dimensional mod MSPE VQ weighting = 0.8 shift by -2 77% sparsity ternary samples
Bits Per Frame	34 (3,4,4,4,4,3,3,3,3,3)	index: 8+6+8+6 ±gain: 5x4	index: 9x4 ±gain: 5x4
Rate	1,133.33 bits/s	1,600 bits/s	1,866.67 bits/s
Miscellaneous	The remaining 200 bits/s are used as follows: 1 bit per frame for synchronization, 4 bits per frame for forward error correction, and 1 bit per frame for future expansion.		

Table 2. CELP Coder Characteristics

4. CODE BOOK SEARCH METHODS

The search procedures for the stochastic and adaptive code books are virtually identical, differing only in their code books and target vectors. To reduce computation, a sequential two-stage search of the code books is performed. The target for the first-stage adaptive code book search is the weighted linear prediction residual plus encoding errors introduced in previous frames that affect the present frame. The second-stage stochastic code book search target is the first stage target minus the filtered adaptive code book VQ excitation.

Let the $L = 60$ dimensional column vectors $\mathbf{s}$, $\hat{\mathbf{s}}$, and $\mathbf{e}$ represent the original speech signal, the synthetic speech signal, and the weighted error signal, respectively. Let $\mathbf{v}$ represent the excitation vector being searched for in the present stage and let $\mathbf{u}$ be the excitation vector of the previous stage. The excitation sequence for a code book of size N within a subframe of size L is characterized by a code book index i, $1 \leq i \leq N$, and a corresponding optimized gain parameter g_i. The excitation vector $\mathbf{v}^{(i)}$ can be written as:

$$\mathbf{v}^{(i)} = g_i \, \mathbf{x}^{(i)}, \tag{10}$$

where the superscript denotes the code book index of the code book vector $\mathbf{x}^{(i)}$.

Let $\mathbf{H}$ and $\mathbf{W}$ be $L \times L$ lower triangular matrices whose columns contain the truncated impulse response of the LP filter and error weighting filter, respectively, excited by a unit impulse on the diagonal.

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ h_{L-1} & \cdots & h_1 & h_0 \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w_0 & 0 & 0 & 0 \\ w_1 & w_0 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ w_{L-1} & \cdots & w_1 & w_0 \end{bmatrix} \tag{11}$$

As shown in Fig. 3, the synthetic speech can be expressed as the convolution of the LP filter's impulse response with its excitation plus its zero input response, $\hat{\mathbf{s}}^{(0)}$:

$$\hat{\mathbf{s}}^{(i)} = \mathbf{H}(\mathbf{u} + \mathbf{v}^{(i)}) + \hat{\mathbf{s}}^{(0)}, \qquad 1 \le i \le N \tag{12}$$

where $\mathbf{u}$ is a zero vector in the first stage search or the scaled adaptive excitation vector in the second stage search.

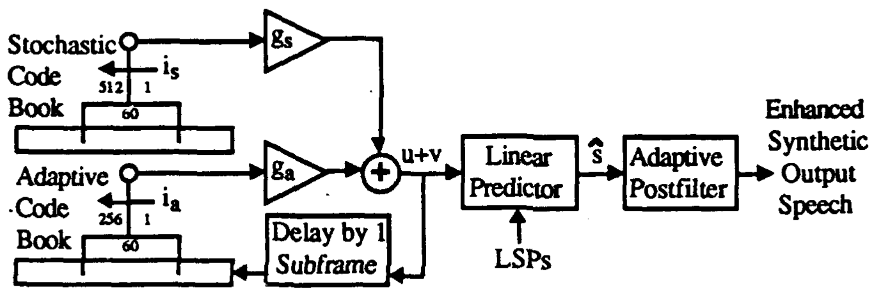


Fig. 3. CELP Synthesizer

As shown in Fig. 4, the weighted error signal is:

$$\mathbf{e}^{(i)} = \mathbf{W}(\mathbf{s} - \hat{\mathbf{s}}^{(i)}) \tag{13a}$$

$$= \mathbf{e}^{(0)} - \mathbf{WHv}^{(i)}, \tag{13b}$$

where the target is:

$$\mathbf{e}^{(0)} = \mathbf{W}(\mathbf{s} - \hat{\mathbf{s}}^{(0)}) - \mathbf{WHu}. \tag{14}$$

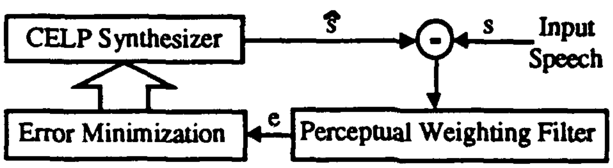


Fig. 4. CELP Analyzer

Thus, the weighted error for codeword i, $e^{(i)}$, is the target minus the scaled filtered codeword:

$$e^{(i)} = e^{(0)} - g_i y^{(i)}, \tag{15}$$

where $y^{(i)}$ represents the filtered codeword:

$$y^{(i)} = \mathbf{WH}x^{(i)}. \tag{16}$$

Let E_i represent the norm or total squared error for codeword i:

$$E_i = \|e^{(i)}\|^2 = < e^{(i)}, e^{(i)} > = e^{(i)T}e^{(i)} \tag{17a}$$

$$= e^{(0)T}e^{(0)} - 2g_i y^{(i)T}e^{(0)} + g_i^2 y^{(i)T}y^{(i)}, \tag{17b}$$

where T denotes transpose. E_i is a function of both the gain factor g_i and the index i. For a given value of i, the optimal gain can be computed by setting the derivative of E_i with respect to the unknown gain value to zero:

$$\frac{\partial E_i}{\partial g_i} = -2y^{(i)T}e^{(0)} + 2g_i y^{(i)T}y^{(i)} = 0. \tag{18}$$

Therefore, the minimum mean squared error gain is the ratio of the cross-correlation of the target and filtered codeword to the energy of the filtered codeword:

$$g_i = \frac{y^{(i)T}e^{(0)}}{y^{(i)T}y^{(i)}}. \tag{19}$$

Minimizing E_i with respect to i is equivalent to maximizing the negative of the last two terms in Eq. (17b) because the first term, $e^{(0)T}e^{(0)}$, is independent of the codeword i. This corresponds to maximizing the match score:

$$m_i = g_i(2y^{(i)T}e^{(0)} - g_i y^{(i)T}y^{(i)}). \tag{20}$$

Substituting Eq.(19) into Eq.(20) yields the familiar normalized squared cross-correlation solution:

$$m_i = \frac{(y^{(i)T}e^{(0)})^2}{y^{(i)T}y^{(i)}}. \tag{21}$$

Thus, as shown in Eq. (21) and Fig. 5, the code book search procedure finds the codeword, i, that maximizes the match score, m_i. This codeword points in the direction closest to the target in the 60 dimensional perceptually weighted space. The magnitude of this vector is determined by a gain factor. Using the above gain term, g_i, results in the minimum squared perceptually weighted error.

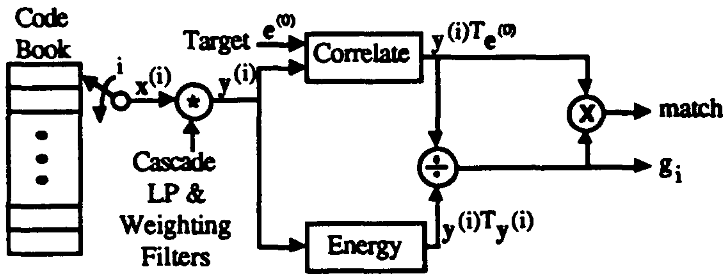


Fig. 5. Code Book Search

While this procedure minimizes mean square error in our model of perceptual space, our listening tests reveal that the subjective speech quality can be improved by modifying the magnitude of the stochastic excitation vector. We introduce this method as *modified excitation*. This method expands upon Shoham's constrained excitation [12] and is compliant with Fed Std 1016. The modified excitation method adaptively attenuates the stochastic code book gain when the long-term predictor is efficient. This increases the relative adaptive code book excitation component by attenuating the stochastic code book excitation component. The CELP coder's subjective speech quality is improved because this reduces roughness and quantizing noise in sustained voiced segments. When the long-term prediction is inefficient, the stochastic code book gain is increased. This provides a more subjectively pleasing match between the unvoiced speech segments of the input and synthesized speech. The efficiency of the long term predictor can be measured by the closeness, in the square-root cross-correlation sense, of the target vectors before and after pitch prediction. Using the previous notation, *R* represents the normalized crosscorrelation and g' represents the modified stochastic code book gain.

$$R = \frac{\left\langle W\left(s-\hat{s}^{(0)}\right),\ W\left(s-\hat{s}^{(0)}\right)-WHu\right\rangle}{\left\|W\left(s-\hat{s}^{(0)}\right)\right\|} \tag{22}$$

$$g' = \begin{cases} 0.2g_i, & |R| < 0.04 \\ 1.4g_i\sqrt{|R|}, & |R| > 0.81 \\ g_i\sqrt{|R|}, & \text{otherwise.} \end{cases} \tag{23}$$

Thus, the modified stochastic excitation is characterized by index i and gain g'_i. Based on Eq. (23), Fig. 6 shows how the gain used in a conventional CELP coder, g_i, is scaled as a function of the efficiency of the long term predictor to yield g'_i. This mapping is the empirical result of our listening tests. Other mappings may result in further enhancements to the speech quality of Fed Std 1016-compatible CELP coders. When this modified excitation technique is performed outside the search loop, it's impact on computation is negligible because the targets can be saved from the searches.

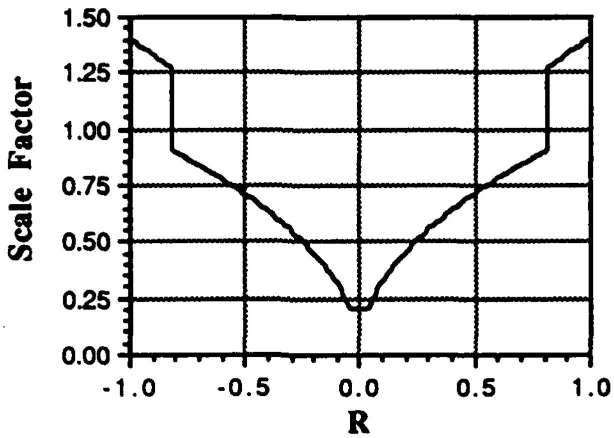


Fig. 6. Modified Excitation Gain Scaling

5. IMPLEMENTATION AND COMPUTATIONAL ESTIMATES

The CELP coder's computations, excluding the code book searches and including the receiver but excluding the code book searches, require approximately 2 million instructions (multiply, add, multiply-accumulate, or compare) per second (MIPS) [13]. The major computational requirements of CELP coding are dominated by the transmitter's code book searches. For a specific implementation to achieve its highest speech quality, trade-offs must be made within each code book search and between the two code book searches. The complexity of the code book search procedures can vary tremendously depending on the techniques used.

To conserve computation, a sequential two-stage search of the code books is performed. Our adaptive code book search consumes approximately 2.3 MIPS per frame [13]. Our stochastic code book search procedure requires approximately 0.0270 MIPS for the first codeword plus 0.00399 MIPS for each additional codeword per subframe [13]. Therefore, to search the whole 512 size stochastic code book four times per frame requires approximately 8.3 MIPS. As shown in Table 3, using our current search procedures, the total upper bound CELP transmitter and receiver computation estimate is 12.6 MIPS. This is an upper bound because we are not claiming to know the fastest code book search procedure. Alternate search domains (e.g., autocorrelation) and code book transformations may yield faster methods [14].

These MIPS estimates should not be confused with DSP chip peak MIPS ratings. Today's state-of-the-art DSP chips require approximately twice the estimated MIPS values, as shown in the last column of Table 3, because of overhead and breaks in the multiply-accumulate pipeline. Although only power-of-2 size code book searches are shown, to achieve the highest quality speech, we expect designers to optimize their implementations to search as many codewords as possible without exceeding their real-time margin.

Stochastic CB Size	Stochastic Search	Total (full-duplex)	DSP Chip Rating
64	1.1 MIPS	5.5 MIPS	11 MIPS
128	2.1 MIPS	6.5 MIPS	13 MIPS
256	4.2 MIPS	8.5 MIPS	17 MIPS
512	8.3 MIPS	12.6 MIPS	25 MIPS

Table 3. CELP Computational Complexity

The proof of these estimates lies in real-time implementation. Many firms have implemented Fed Std 1016 coders, including Analog Devices, AT&T, DSP Software Engineering, GE/RCA, Intellibit, Motorola, Technical Communications Corp., and Titan Linkabit. Real-time full-duplex Fed Std 1016 coders have been demonstrated based on a single DSP chip (e.g., Texas Instruments' 16.6 MIPS floating-point TMS320C31 or Motorola's 13.3 MIPS integer DSP56001). A high-quality, low-power, small-sized voice processor can be constructed for under $200 parts cost in small quantities by adding to one of these DSP chips: ROM, 16k words of SRAM, and a Texas Insturments TLC32044 A/D and D/A with filters chip. The 'C31 and '56001 CELP implementations provide high quality speech by searching half the stochastic code book and hierarchically searching the adaptive code book's integer delays followed by neighboring noninteger delays. This confirms our 17 MIPS DSP chip estimate and demonstrates the feasibility of high-quality and low-cost Fed Std 1016 implementations.

6. PERFORMANCE

Digital voice coding can offer significant advantages over conventional analog voice. Digital coders are impervious to channel noise below the error threshold of the modem. A striking example of this is in land mobile radio (LMR) channels where noise can severely degrade analog communications [4]. The low data rate of the CELP coder allows strong error protection over the harsh, narrow LMR channel for high-quality voice communications that would otherwise be infeasible.

Measuring the voice performance of the low-rate speech coders is difficult because the quest for an objective voice performance metric has been elusive. Low-rate coders do not match the input waveform, so conventional objective metrics, such as signal-to-noise ratio (SNR), are inappropriate and can be misleading. For example, a CELP coder can have a lower SNR than a subjectively inferior CVSD coder because the CELP coder matches a perceptual criterion rather than the waveform. Thus, low-rate coders are best assessed by subjective means; i.e., a listening panel. Subjective performance can be analyzed on the basis of intelligibility or quality. Intelligibility is determined by how well people can communicate with each other. Quality is determined from the fidelity of the voice.

Voice performance assessment is further complicated by the effects of acoustic background noise and channel errors. It is crucial for a speech coder to behave well (i.e., degrade gracefully) in these real-world conditions. Speech coders that do well in quiet environments and with clear channels may be too fragile for the real-world conditions of a given application. For example, speech coders with voicing detectors and pitch trackers are often susceptible to

falsely tracking on background noise instead of the speaker's voice. Thus, speech coders must be assessed by listeners in a variety of conditions appropriate to a given application.

Fig. 7 shows subjective mean opinion scores (MOS) for 2,400 bit/s LPC-10e, 32,000 bit/s ADPCM, 64,000 bit/s μlaw PCM and CELP coders at 4,800, 8,000 and 16,000 bits/s. In MOS testing, listeners (using telephone handsets) rate speech coders using the subjective labels shown on the vertical axis. CELP coding at 16,000 bits/s offers subjective performance equivalent to 32,000 bit/s ADPCM. Although Fed Std 1016 4,800 bit/s CELP coding is below this performance level, it still provides good performance.

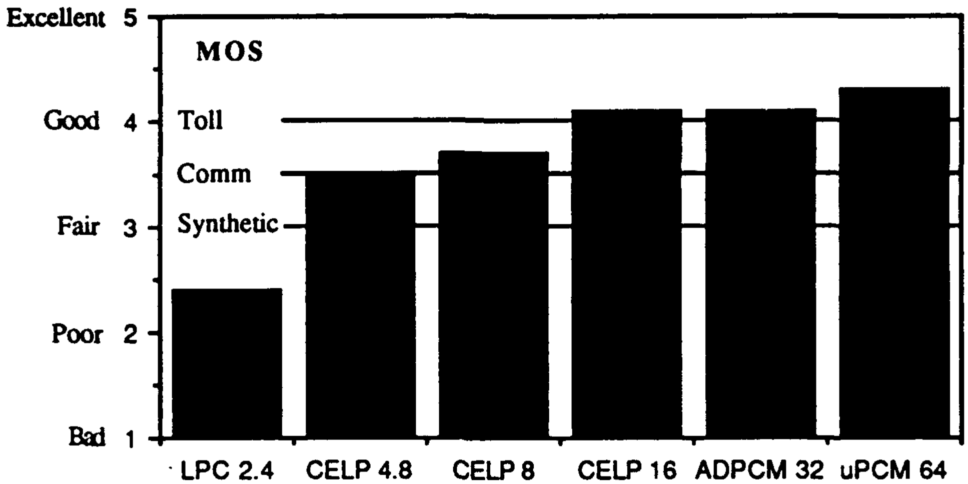


Fig. 7. MOS speech coder performance

We formally measure speech intelligibility and quality subjectively using Dynastat's diagnostic rhyme test (DRT) and diagnostic acceptability measure (DAM). Typical test variances are one point for DRT and two points for DAM scores. Our Fed Std 1016 implementation scores 93% intelligibility on the DRT test [3].

Figures 8 and 9 show subjective quality scores evaluated in different environments for input speech (whose quiet DAM is off the scale at 84), long-distance plain old telephone service (POTS), and common narrowband U.S. Government standard coders. Coder performance is shown in Fig. 8 for quiet, office, E-3A/E-4B Airborne Command Post compartment environments, and a 1% uniform random bit error rate condition. In Fig. 9, tandem coding of each coder into 2,400 bit/s LPC-10e is represented by ->LPC and 16,000 bit/s CVSD into each coder is shown by CVSD->.

High frequency response is crucial to DAM quality scores. POTS (with a 3,500-Hz cutoff) and CVSD 16 (with a 3,000-Hz cutoff) score low relative to the digital coders with a 3,800-Hz cutoff such as CELP. LPC vocoders have an unnatural synthetic quality which lowers their speech quality scores. Two important factors affecting performance are acoustic background noise and tandem coding. Acoustic background noise (e.g., office noise) degrades all the coders and, as shown in Fig. 9, it has an equalizing effect on the DAM scores between coders. CELP coders do not exhibit the usual vocoder problems in background noise because they use a more sophisticated excitation model than the classical vocoder's pitch and voicing (e.g., LPC-10). Background noise, including multiple speakers, is faithfully reproduced. As shown in the figures, CELP coding performance is outstanding among other coders. CELP coding, at 4,800 bits/s, breaks the performance barrier of most Government standards, providing Consortium ratings of "very good" intelligibility and "excellent" quality, comparable to 32,000 bit/s CVSD. CELP coding will usher in a new era of narrowband speech coders capable of receiving wide user acceptance by providing very high quality speech.

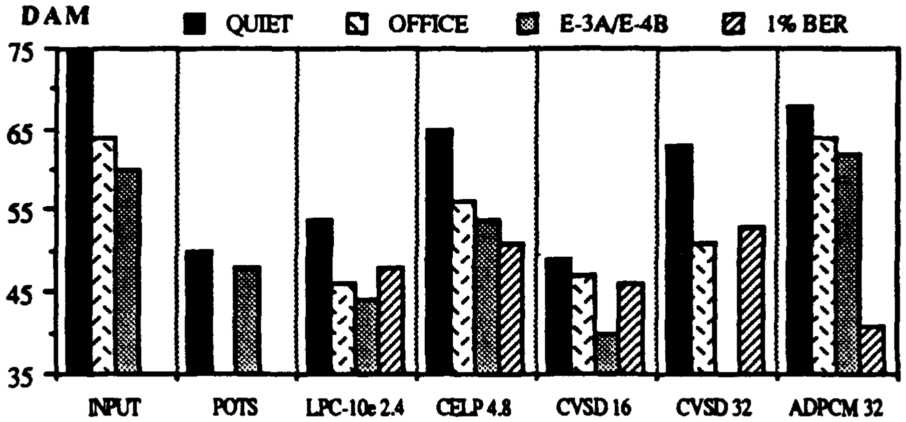


Fig. 8. Government standard speech coder quality comparison

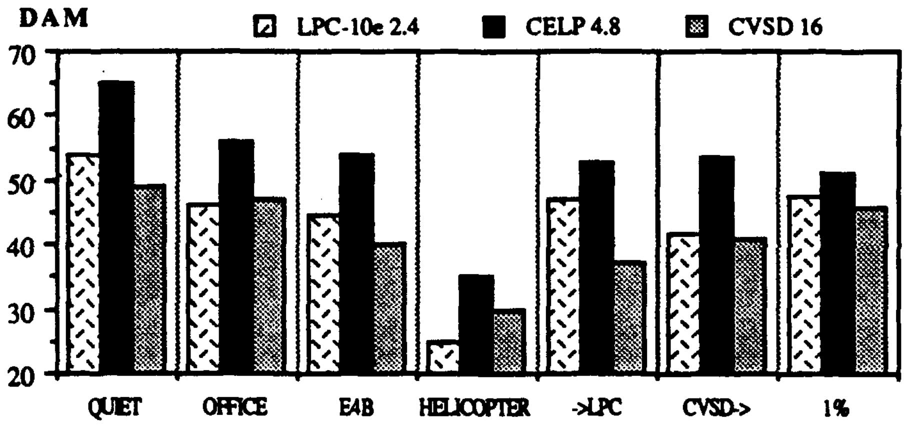


Fig. 9. Speech coder quality for LPC-10e, CELP and CVSD 16

7. CONCLUSIONS

Fed Std 1016 based CELP coding is robust in real world conditions (noisy environments, nonspeech input, tandem coding, and transmission errors). This standard provides the flexibility to allow for computation reductions and performance enhancements based upon new search procedures and more sophisticated perceptual models. Most importantly, Fed Std 1016 is practical to implement and sets an expandable 4,800 bit/s coding standard that provides high quality speech for many applications, including secure voice, portable telephones, and Land Mobile Radio.

II. IMPLEMENTATION DISCUSSION

1. CELP Details

1.1 Input and Output Conditioning

1.1.1 Filtering. The CELP coder input passband should be essentially flat from 200 to 3,600 Hz. A typical input filter has 3 dB attenuation points at 100 and 3,800 Hz; less than 1 dB of inband ripple; and minimum attenuations of 18 dB at 50 Hz, 18 dB at 4,000 Hz, and 46 dB above 4,400 Hz. A template of the filter is shown in Fig. 10. The output filter is similar to the input filter with the addition of any digital-to-analog compensation needed for flat response (e.g., sin(x)/x). We suggest using a 200 Hz 2nd order Butterworth high pass output filter.

1.1.2 A/D Conversion. Analog-to-digital conversion shall use an 8 kHz $\pm$ 0.1 percent sampling frequency and have a dynamic range of at least 12 bits.

1.2 CELP Analyzer

CELP uses an 8 kHz sampling rate and a 30 ms frame size divided into four 7.5 ms subframes. CELP analysis consists of three basic functions: 1) short delay "spectrum" prediction, 2) long delay "pitch" search, and 3) innovation "code book" search. CELP synthesis consists of the same three functions (performed in reverse order) with the addition of a fourth function, called a postfilter, to enhance the output speech. The CELP encoder's characteristics are summarized in Table 2 (Section I) and are briefly described in the following. As shown in Fig. 11, the CELP encoder consists of the following four branches:

1.2.1 Spectrum Analysis. Spectrum analysis parameters are transmitted once per frame. Recommended spectral analysis techniques are given in Section 2.2. The spectrum is coded using 34 bit independent nonuniform scalar quantization of line spectral parameters (LSP). A spectral parameter interpolation scheme, as given in Section 2.2, is recommended because spectral parameters are transmitted only once per frame, but they are needed for each of the 4 subframes.

The past spectrum and future spectrum are centered at the beginning and end of the present frame's excitation parameters, respectively. This spectrum alignment means that the spectrum is computed one-half a frame (2 subframes) ahead of the excitation parameters.

1.2.2 Adaptive Code Book Search. Pitch search is performed using an adaptive code book gain shape VQ method. The adaptive code book structure is shown in Table 4. The allowable pitch delays are given in Section 1.4.2. The pitch delay ranges from 20 to 147 lags (including noninteger values) every odd subframe, while even subframes are coded within 64 indices relative to the previous subframe. Recommended search procedures are given in Section 2.

The pitch search is performed and coded four times per frame (every 7.5 ms). The pitch gain is coded between -1 and +2 using 5 bit absolute nonuniform scalar quantization.

1.2.3 Stochastic Code Book Search. Stochastic code book search is performed by a fixed code book gain shape VQ method. The stochastic code book structure is shown in Table 5. The allowable code book indices range from 0 to 511. Recommended search procedures are given in Section 2.

The stochastic code book search is coded four times per frame (every 7.5 ms). The code book gain is coded using 5 bit absolute nonuniform scalar quantization as shown in Table 6.

1.2.4 Update. The analyzer's model of the synthesizer is run to update all the filter states, which are then used by the pitch and code book searches in the next frame of speech.

1.2.5 Analyzer Software Flowchart. The analysis subroutine hierarchy is shown in Fig. 12.

1.3 CELP Synthesizer

The CELP synthesizer is shown in Fig. 1 (Section I). The parameters required are the code book index and gain, the pitch delay and gain, and the spectrum predictor parameters. A postfilter can be added which may enhance the synthesized speech. However, in noisy environments or tandem coding situations, postfiltering can be detrimental if not used with caution.

1.4 Coding and Decoding

1.4.1 Spectrum. The 10 line spectral parameters (LSP) shall be coded with the number of bits per parameter specified in Table 7. The encoding is by a nearest output level monotonically constrained scalar quantizer. The encoding/decoding tables for the 10 LSPs is given in Table 8.

1.4.2 Adaptive Code Book. The pitch delay shall be coded with the number of bits, as a function of the subframe, specified in Table 7. In subframes 1 and 3, the pitch delay can be any value given in Table 9. In subframes 2 and 4, the pitch delay shall be delta coded relative to the index (range 0 to 255) of the previous subframe's delay shown in Table 9. The delta coding range shall be between the delay values indexed by MIN and MAX inclusive as follows:

let MIN = index of previous delay - 31; if MIN < 0: MIN = 0, MAX = 63
let MAX= index of previous delay + 32; if MAX> 255: MIN = 192, MAX = 255

For example, to repeat a previous delay between 29.5 and 114, the 6-bit delta code is 011111.

The pitch gain shall be coded with the number of bits shown in Table 7. The coding/decoding is a nearest output level scalar quantizer. The encoding/decoding table for the pitch gain is given in Table 10.

1.4.3 Stochastic Code Book. The stochastic code book is formed by extracting overlapping samples from a code vector to form each codeword. The code book samples are ternary valued (-1, 0, or +1). The code book is overlapped by a shift of 2 samples to allow computational savings by recursive end-point correction algorithms in the code book search. Code book samples are shown in Table 11. The structure of the code book is given in Table 5. The code book index shall be coded with the number of bits shown in Table 7.

The code book gain shall be coded with the number of bits shown in Table 7. The coding/decoding is a nearest output level scalar quantizer. The encoding/decoding table for the code book gain is given in Table 6.

1.4.4 Synthesizer Software Flowchart. The synthesis subroutine hierarchy is shown in Fig. 13.

1.5 Error Protection

1.5.1 Overview. The primary goal of error protection is the prevention of perceptually disturbing synthesis errors: loud clipped speech (blasts) and squeaks. Parameter coding and continuity are the basis for the error protection strategy. Application of adaptive smoothers as described in Section 2.12 are recommended.

Forward error correction is performed by a Hamming (15,11) single error correcting and detecting code that protects 10 pitch delay and pitch gain bits. The pitch delay bits PD(1)-5, 6, 7 and PD(3)-5, 6, 7 (as defined in Table 12) of the absolute pitch delay are protected. The most significant bits of the pitch gain, PG(1)-4, PG(2)-4, PG(3)-4 and PG(4)-4, are protected. For future expansion, 1 bit per frame is reserved. This is the 11th bit protected by the Hamming code.

1.5.2 Description. Four bits shall be allocated to forward error correction coding using a Hamming (15,11) code. Tables 13 and 14 show the data bits to be protected and the decoding table.

The data bits to be protected are ordered as shown in Fig. 14, and a parity check bit is generated for each of the four fields shown. Check bits 1, 2, 3 and 4 are set to 1 if the data bits in their respective fields have odd parity (i.e., the code has even parity). In the decoding process, the procedure is repeated, and the calculated check bits are EXCLUSIVE OR'ed

(XOR) with the received check bits. The 4 bit result of the XOR is used as an index to Table 12. This table shows which bit position in the data word of Fig. 14 to invert.

1.6 Transmission Format

1.6.1 Transmission Rate. The transmission rate shall be 4,800 bits/s ± 0.01 percent. All frames contain 144 bits. The spectrum frame length is 30 ms ± 0.01 percent. The pitch and code book subframe length is 7.5 ms.

1.6.2 Bit Allocation. The allocation of the 144 bits in a CELP frame shall be as shown in Table 5.

1.6.3 Bit Assignment. The assignment of bits within a CELP frame shall be as shown in Table 10. In frame 1, the spectrum for frame 1 along with default excitation (pitch and code book) parameters for the first 2 subframes of frame 1 and analyzed parameters for the following 2 subframes of frame 1 are transmitted. Likewise, successive frames are transmitted with the spectrum parameters 2 subframes ahead of the excitation parameters.

The transmitted bit stream is formed by assembling the CELP parameter symbols (code book indices and gains and the linear prediction coefficients), the Hamming code, and future expansion bit. Except for the odd subframe adaptive code book index, all the symbols are formed by natural binary coding (NBC) of the quantized parameters. The NBC assigns all ZEROS to the smallest value and increments by unity to all ONES for the largest value. The symbols for the odd subframe adaptive code book index are given in Table 7. The symbols are then assembled in the bit stream in the order shown in Table 10.

1.6.4 Synchronization. The synchronization bit shall alternate between ZERO and ONE from frame to frame. The first transmitted frame shall start with ZERO.

1.6.5 Spare. The spare bit shall be set to ZERO. The spare bit will allow for future upgrades to the coder.

2. Comments and Recommendations

The following list of comments and recommendations may be beyond the scope of interoperability with Fed Std 1016.

2.1 Scaling

The scaling of the input speech, the impulse, and the stochastic code book samples and gains are all interrelated. The tables given here assume that the input speech has a range of ± 32,767.0 and the impulse response of the filters is calculated using 1.0 for the impulse.

2.2 Spectrum Analysis.

Spectrum analysis is performed once per frame by open-loop, 10th order autocorrelation LPC analysis using no preemphasis and a 15 Hz bandwidth expansion using a 30 ms Hamming window. The spectrum is coded using 34 bit independent nonuniform scalar quantization of line spectral parameters (LSP). The LSPs are linearly interpolated to form an intermediate set for each of the four subframes. The interpolation weights are:

Subframe	Past Spectrum	Future Spectrum
1	7/8	1/8
2	5/8	3/8
3	3/8	5/8
4	1/8	7/8

2.3 Search Procedures

Using quantized and interpolated spectral parameters, the excitation parameters are determined sequentially for each subframe. After selecting the adaptive code book index and gain, the stochastic code book index and gain are determined.

Use of exhaustive analysis-by-synthesis search procedures is not required. Nonexhaustive or suboptimal search procedures can be interoperable with this standard. The

pitch search is especially ripe for nonexhaustive search procedures. The code book search procedures are derived in [15] and [16].

2.4 Adaptive Code Book Search.

Pitch search is performed by closed-loop analysis using a modification of what is commonly called any one of the following: self-excited, adaptive code book, or VQ method. As in Fig. 5 (Section I), the adaptive code book is convolved with the perceptual weighting filter's impulse response. For an exhaustive search, the convolution is calculated for each of the allowable pitch lags. The allowable pitch lags are given in Section 1.4.2. Each pitch lag's convolution is then correlated with the e' short-delay (spectrum only) predictor's speech residual. The pitch delay which minimizes the squared prediction error (MSPE) corresponds to the peak in the match score (normalized squared crosscorrelation function). This delay is selected for transmission unless a submultiple delay's squared prediction error is within 1/2 dB of the peak, in which case the submultiple delay is selected. Submultiples of 2, 3 and 4 are tested and the shortest submultiple delay satisfying this criteria is chosen.

2.5 Noninteger Pitch Delay

Noninteger values of pitch delay may be handled without increasing the 8 kHz sampling rate. The adaptive code book may be resampled or polyphase filtered to obtain noninteger pitch delays. We suggest using an 8-point Hamming windowed sinc resampling function in the pitch search loop. For reconstruction we suggest using a more accurate 40-point Hamming windowed sinc resampling function.

2.6 Construction of Integer and Noninteger Pitch Codewords

A brief explanation on the construction of integer and noninteger valued pitch codewords is shown here:

Let **r** represent the adaptive code book stored as a linear array of overlapped codewords:

$$\mathbf{r} = (r(-147), r(-146), \ldots, r(-1)).$$

Let **r'** represent a candidate codeword derived from the adaptive code book:

$$\mathbf{r'} = (r'(0), r'(1), \ldots, r'(59)).$$

Let **r"** represent the concatenation of **r** and **r'**:

$$\mathbf{r''} = (r(-147), r(-146), \ldots, r(-1), r'(0), r'(1), \ldots, r'(59))$$

$$= (r''(-147), r''(-146), \ldots, r''(-1), r''(0), r''(1), \ldots, r''(59)).$$

For an integer delay, M, a codeword, **r'**, is constructed by repeating the previous excitation signal, **r**, delayed by M samples. For delays less than the subframe length (M < 60), the adaptive code book, **r**, contains the initial M samples of the codeword **r'**. To complete the codeword to 60 elements, the short vector is replicated by periodic extension. Thus, an integer delay candidate codeword **r'** at delay M is:

$$r'_M(i) = r''_M(i) = r''_M(i-M), \qquad \text{where } i = 0, 1, \ldots, 59, \quad M = 20, 21, \ldots, 147.$$

Note: For M < 60, the **r"** array is filled recursively and the order of index assignment must be performed in the specified order.

For noninteger delays, the codewords are formed by interpolation. The interpolation used for synthesis in the transmitter and receiver must be interoperable with a 40-point interpolation using the weights, **w**, of the Hamming windowed sinc function:

$$h(k) = 0.54 + 0.46 \cos\left(\frac{\pi k}{6N}\right), \qquad \text{where } k = -6N, -6N+1, ..., 6N$$

$$w_f(j) = h(12(j+f))\,\frac{\sin(\pi(j+f))}{\pi(j+f)}, \qquad \text{where } j = -N/2, -N/2+1, ..., N/2-1$$

$$f = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}$$

where N is an even number of interpolation points. For example, the first few weights of an N = 8 point interpolation for $f = \frac{2}{3}$ are:

$w_{2/3}(-4) = -0.11713e-01$ $\qquad$ $w_{2/3}(-2) = -0.15920e+00$

$w_{2/3}(-3) = 0.49731e-01$ $\qquad$ $w_{2/3}(-1) = 0.81403e+00$

Notes: Instead of a 40-point interpolation, as few as N = 8 points using the above formula can give acceptable performance.
The interpolations used for synthesis in the transmitter and receiver are not required to be identical to the interpolation used in the adaptive code book search.
For example, using an 8-point window for search and a 40-point window for synthesis is a reasonable option.

The noninteger delay consists of an integer part M plus a fractional part f. The fractional part of the delay determines which set of weights is used to form the interpolated codeword. A recursive interpolation formula may be used to calculate the codeword **r'** at delay M+f:

$$r'_{M+f}(i) = r''_{M+f}(i) = \sum_{j=-N/2}^{N/2-1} w_f(j)\, r''_{M+f}(i-M+j), \quad \text{where } i = 0, 1, ..., 59$$

$$M = 20, 21, ..., 147$$

$$f = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}$$

Note: The **r''** array is computed recursively and the order of index assignment must be performed in the specified order.

Finally, after completing the code book searches, the adaptive code book, **r**, is updated with the chosen excitation vector, **e** (the sum of the chosen scaled stochastic and adaptive codewords). The update shifts the code book array and shifts in the excitation vector:

$$r(i) = r(i+60), \qquad \text{where } i = -147, -146, ..., -61$$

Note: The index assignment must be performed in the specified order

$$r(i) = e(i+60), \qquad \text{where } i = -60, -59, ..., -1$$

and where e(0) is the first sample in time used to excite the LP synthesis filter.

An example of noninteger pitch for delays less than the subframe length is shown in Fig. 15.

2.7 Stochastic Code Book Search.

2.7.1 Overview

Stochastic code book search is performed by closed-loop analysis using conventional MSPE criteria of the perceptually weighted error signal. As shown in Fig. 15, the code book is convolved with the perceptual weighting filter's impulse response. The convolution is calculated for a maximum of 512 codewords. Each code word convolution is then correlated with the e'' speech residual from the short-delay (spectrum) and long-delay (pitch) predictors. The codeword selected for transmission maximizes the match score function over the searched code book, resulting in MSPE.

2.7.1 Modified Excitation A simple modification of the gain term, shown in Fig. 15, reduces CELP's quantizing noise.

Depending on the current system state, the stochastic code book excitation is reduced to a level that is low enough to produce positive perceptual effects, yet is high enough so as not to upset the dynamics of the system. The main effect of the method is that during sustained voiced sounds, the excitation level is attenuated. In unvoiced and transition regions the level is amplified to a level slightly more than that of standard CELP.

The relative adaptive code book excitation component is increased in voiced regions by decreasing the stochastic code book excitation component. The amount of decrease in the stochastic component depends on the efficiency of the adaptive component. More reconstruction burden is placed on the adaptive component as its efficiency increases. The efficiency is measured by the closeness (in the squareroot crosscorrelation sense) of the residual signals before and after pitch prediction. When the efficiency is high (e.g., > 0.9), the stochastic component is amplified slightly (e.g., one quantizer level).

The procedure for modifying the stochastic gain outside the search loop is:

1) Measure the efficiency of the adaptive component
2) Search the stochastic code book for the optimum codeword
3) Modify the stochastic code book gain

2.8 Perceptual Weighting

Different perceptual weighting factors may be interoperable with this standard. We use a weighting factor equal to 0.8 in the short-term predictor for both the stochastic and adaptive code book searches. Weighting factors less than unity expand the bandwidths by moving the poles radially in the z-plane toward the origin. For a weighting factor of 0.8 and a sampling rate of 8 kHz, the bandwidth expansion is -(8,000 Hz/pi) ln(0.8) = 568 Hz and the predictor coefficients a_i are scaled by 0.8^i.

2.9 Default Excitation Parameters

Different default excitation parameters (used in the first 2 subframes) may be interoperable with this standard. The present default excitation parameters are set to their coded values closest to ZERO. However, it may be possible to use these defaults for additional purposes (e.g., signaling).

2.10 Filter Structure

Different filter structures may be an interoperability issue. The filter type must be compatible with direct form (block wise) filters. Lattice form (sample-by-sample) filters may have some implementation advantages.

2.11 Postfilter

Cautious application of postfiltering at the synthesizer's output is recommended. The ear's masking properties are exploited to trade off speech distortion vs quantizing noise. In tandem coding scenarios, only one stage of postfiltering is recommended and multiple stages should be avoided.

2.12 Error Protection

The primary goal of error protection is the prevention of perceptually disturbing synthesis errors: loud clipped speech (blasts) and squeaks. Parameter coding and continuity are the basis for the error protection strategy. Adaptive nonlinear smoothers based on estimates of the channel error rate from the forward error correcting (15,11) Hamming coder are recommended. Therefore, the smoothers do not operate in error free conditions.

When clipping occurs in the output subframe, we suggest attenuating all the samples in that subframe by 30 dB to avoid extraneous blasts in the output speech.

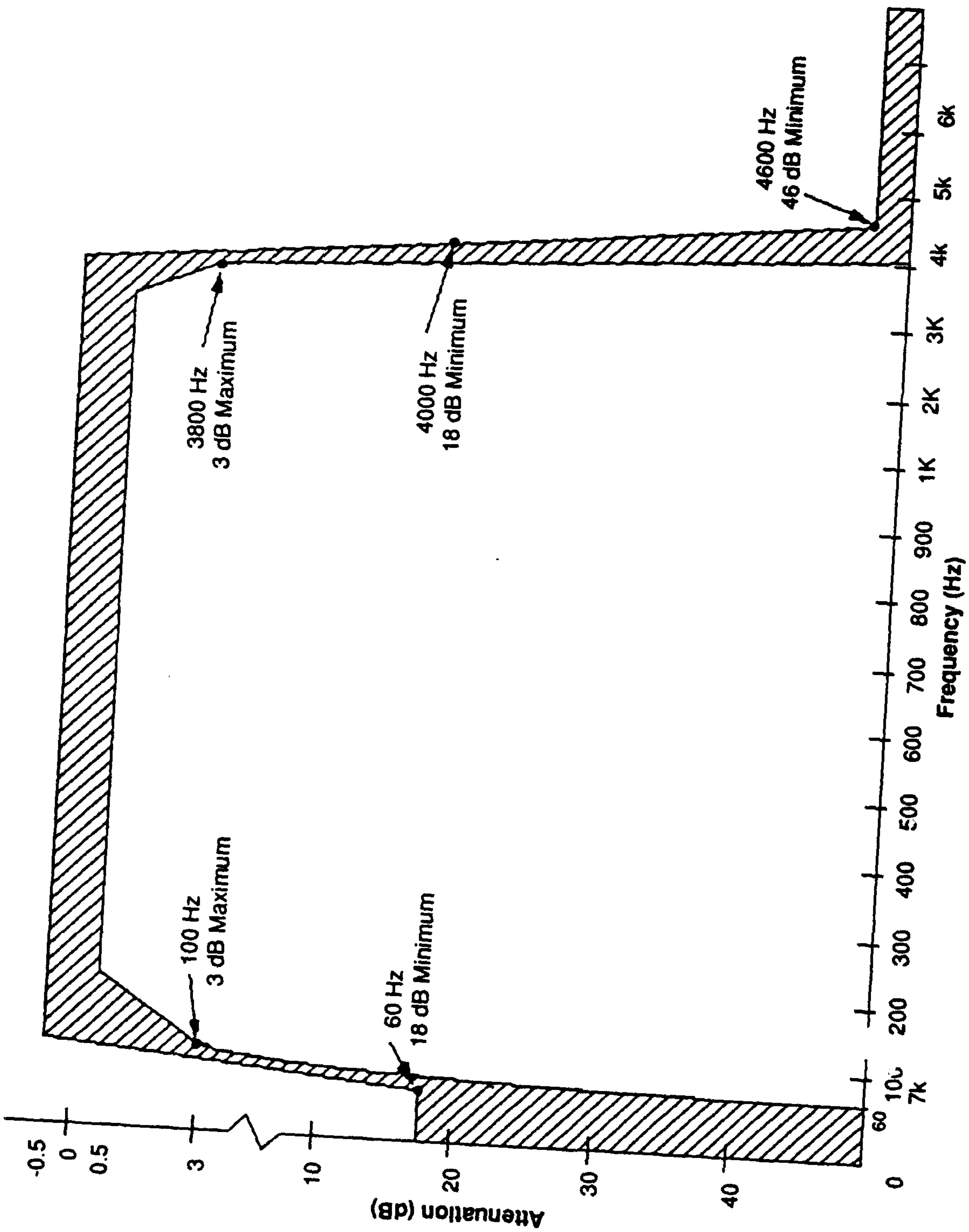


Figure 10. Input Filter Template

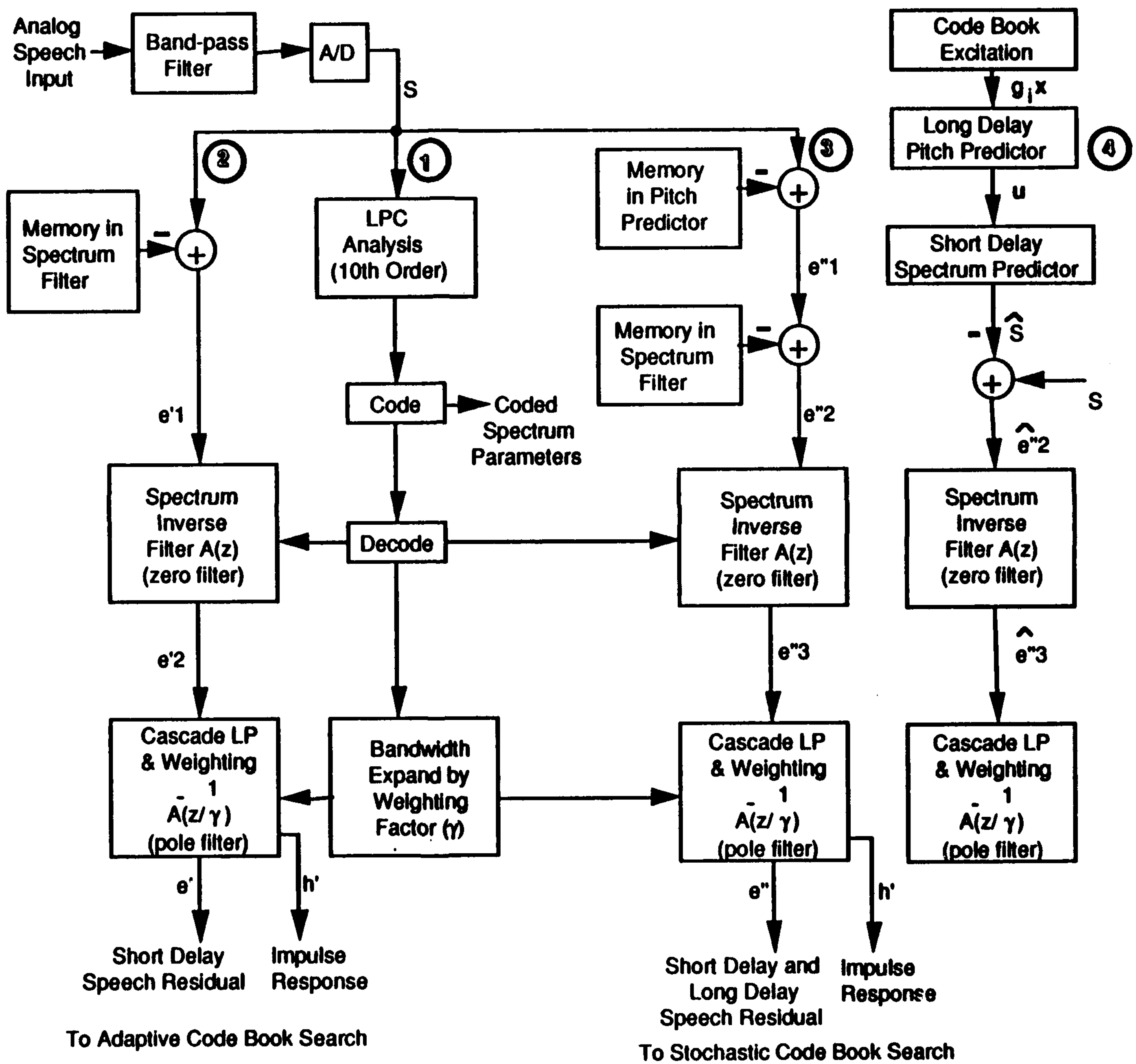


Figure 11. CELP Encoder

Index	Delay	Adaptive Code Book Sample Numbers
255	147	-147, -146, -145, ..., -89, -88
. . .	. . .	. . .
131	61	-61, -60, -59, ..., -2, -1
. .	. .	. .
128	60	-60, -59, -58, ..., -1, 0
. . .	. . .	. . .
3	21	-21, -20, ..., -1, -21, -20, ..., -1, -21, -20, ..., -4
. .	. .	. .
0	20	-20, -19, ..., -1, -20, -19, ..., -1, -20, -19, ..., -1

Table 4. Adaptive Code Book Structure

Index	Stochastic Code Book Sample Numbers
511	0, 1, 2, ..., 58, 59
510	2, 3, 4, ..., 60, 61
. . .	. . .
N	2(511-N), 2(511-N)+1, ..., 2(511-N)+59
. . .	. . .
1	1020, 1021, 1022, ..., 1078, 1079
0	1022, 1023, 1024, ..., 1080, 1081

Table 5. Stochastic Code Book Structure

3.2 CELP Hierarchical Chart

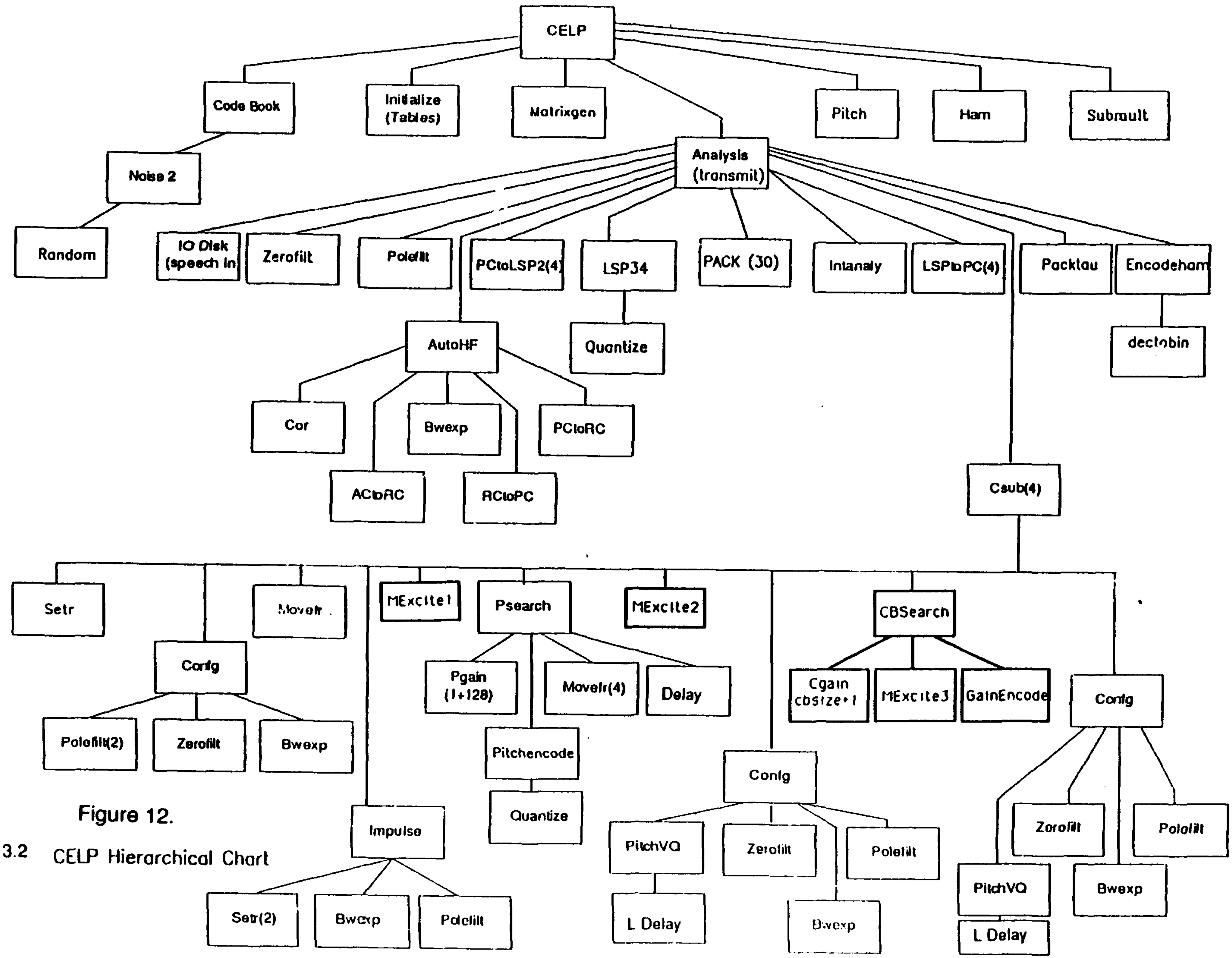


Figure 12.

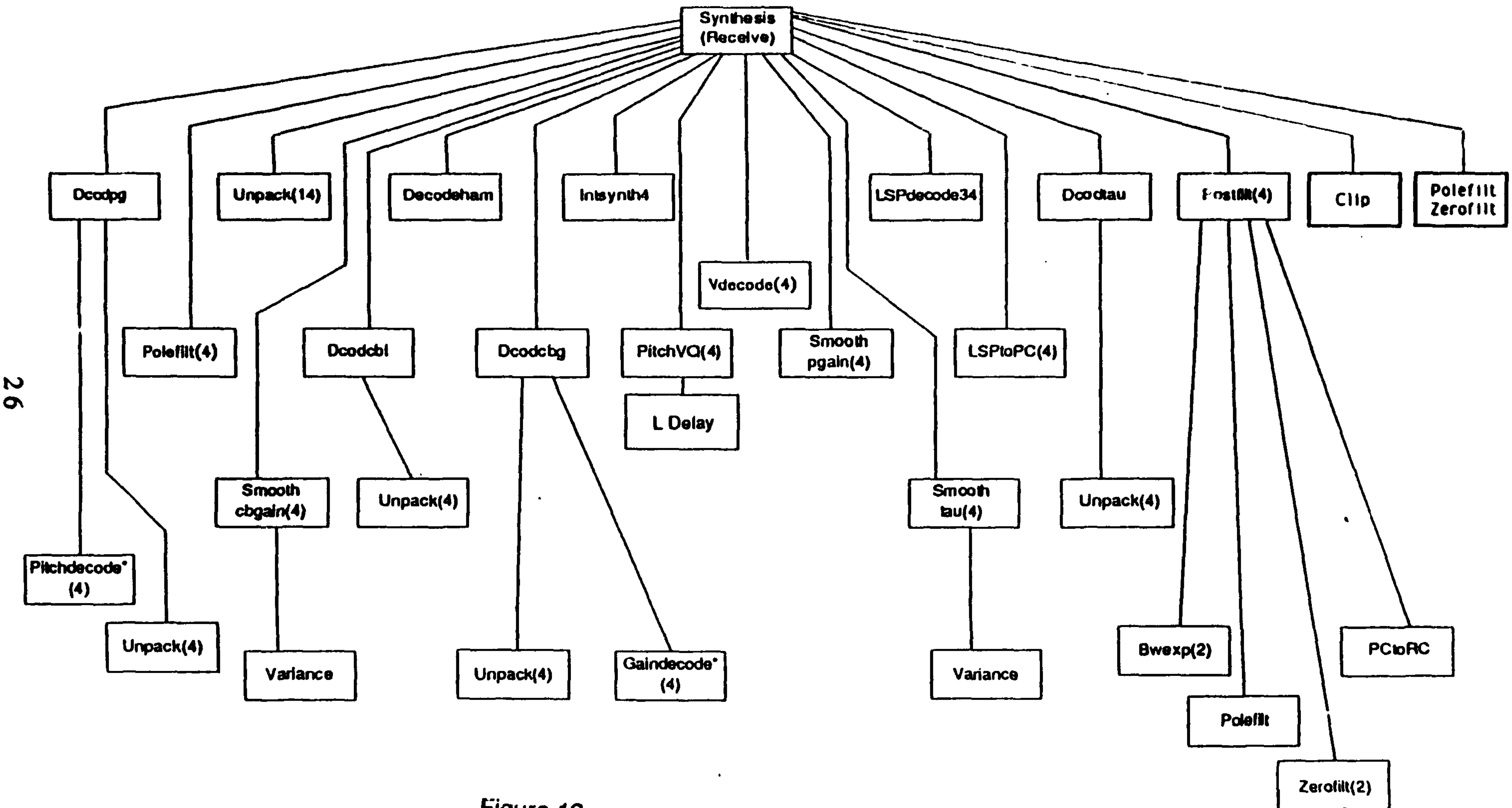


Figure 13.

3.2 CELP Hierarchical Chart

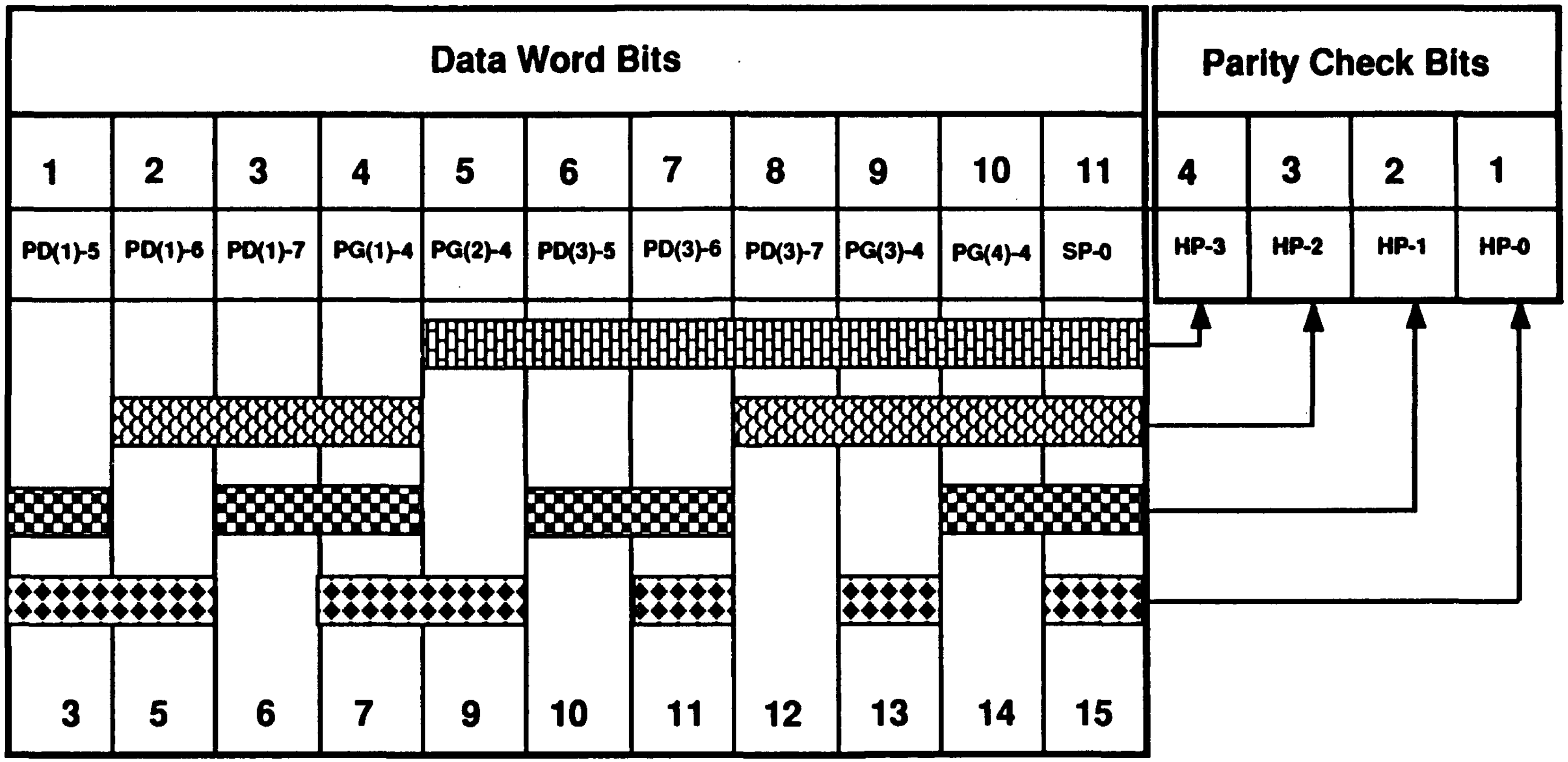
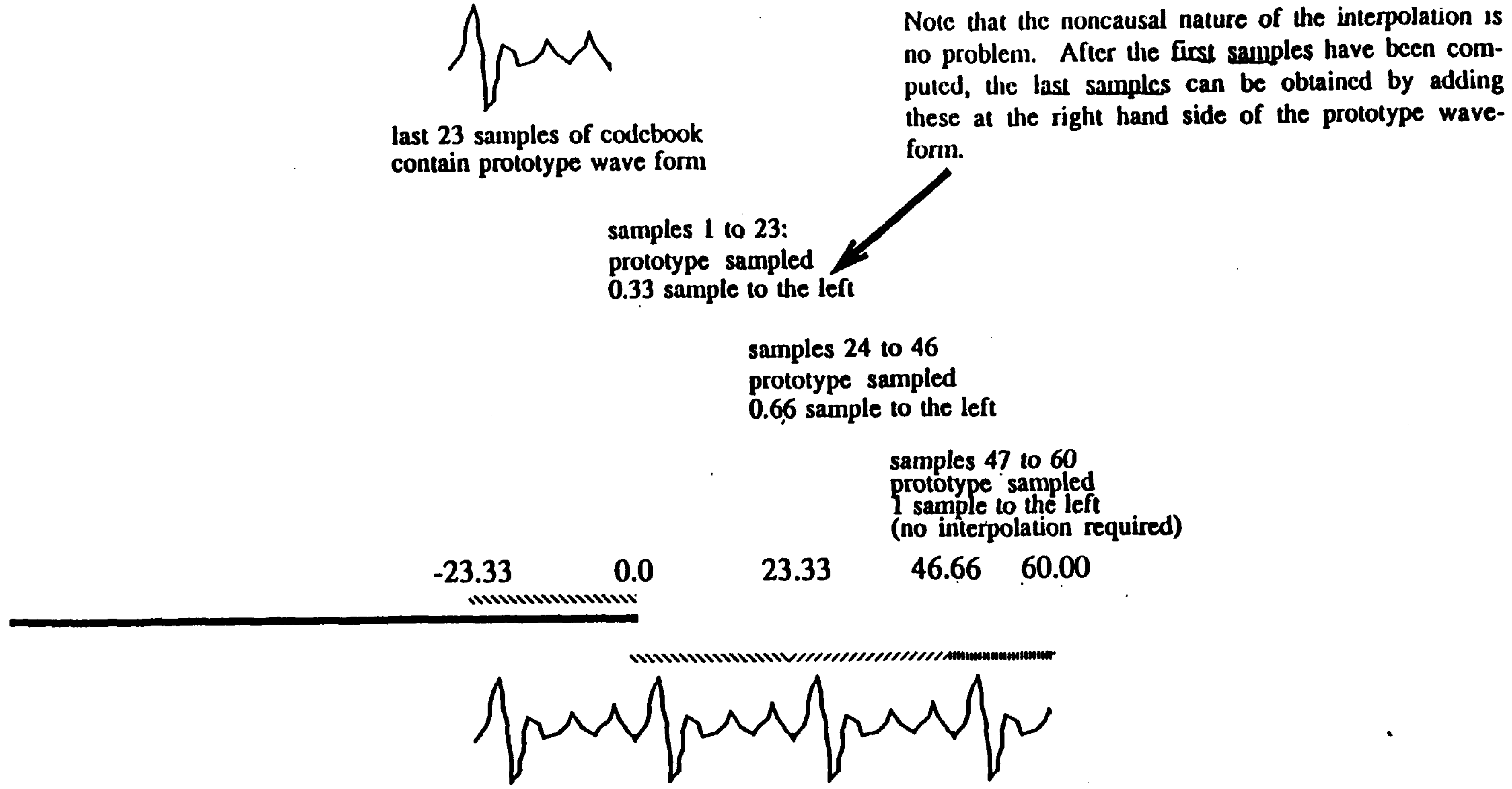


Figure 14. Hamming (15, 11) Foward Error Correction



example of non-integer adaptive codebook search for pitch lag of 23.33 samples

Figure 15.

-1330,	-870,	-660,	-520,	-418,	-340,	-278,	-224,
-178,	-136,	-98,	-64,	-35,	-13,	-3,	-1,
1,	3,	13,	35,	64,	98,	136,	178,
224,	278,	340,	418,	520,	660,	870,	1330

Table 6. Code Book Gain Encoding/Decoding Levels

Parameter	Subframe				Frame
	1	2	3	4	
LSP 1					3
LSP 2					4
LSP 3					4
LSP 4					4
LSP 5					4
LSP 6					3
LSP 7					3
LSP 8					3
LSP 9					3
LSP 10					3
Pitch Delay	8	6	8	6	28
Pitch Gain	5	5	5	5	20
Code Book Index	9	9	9	9	36
Code Book Gain	5	5	5	5	20
Future Expansion					1
Hamming Parity					4
Synchronization					1
Total					144

Note: All parameters are coded LSB to MSB

Table 7. Bit Allocation

LSP	Bits	Output Levels (Hz)
1	3	100, 170, 225, 250, 280, 340, 420, 500
2	4	210, 235, 265, 295, 325, 360, 400, 440, 480, 520, 560, 610, 670, 740, 810, 880
3	4	420, 460, 500, 540, [illegible]85, 640, 705, 775, 850, 950, 1050, 1[illegible]0, 1250, 1350, 1450, 1550
4	4	620, 660, 720, 795, 880, 970, 1080, 1170, 1270, 1370, 1470, 1570, 1670, 1770, 1870, 1970
5	4	1000, 1050, 1130, 1210, 1285, 1350, 1430, 1510, 1590, 1670, 1750, 1850, 1950, 2050, 2150, 2250
6	3	1470, 1570, 1690, 1830, 2000, 2200, 2400, 2600
7	3	1800, 1880, 1960, 2100, 2300, 2480, 2700, 2900
8	3	2225, 2400, 2525, 2650, 2800, 2950, 3150, 3350
9	3	2760, 2880, 3000, 3100, 3200, 3310, 3430, 3550
10	3	3190, 3270, 3350, 3420, 3490, 3590, 3710, 3830

Table 8. Spectrum Encoding/Decoding Levels

-0.993,	-0.831,	-0.693,	-0.555,	-0.414,	-0.229,	0.000,	0.139,
0.255,	0.368,	0.457,	0.531,	0.601,	0.653,	0.702,	0.745,
0.780,	0.816,	0.850,	0.881,	0.915,	0.948,	0.983,	1.020,
1.062,	1.117,	1.193,	1.289,	1.394,	1.540,	1.765,	1.991

Table 10. Pitch Gain Encoding/Decoding Levels

20.00	42	34.67	C0	51.67	98	68.67	C5	97.00	A1
20.33	46	35.00	C3	52.00	90	69.00	C9	98.00	97
20.67	47	35.33	C2	52.33	80	69.33	C8	99.00	87
21.00	57	35.67	D2	52.67	9A	69.67	C7	100.00	9F
21.33	56	36.00	D3	53.00	8A	70.00	CB	101.00	8F
21.67	59	36.33	D1	53.33	82	70.33	C6	102.00	81
22.00	58	36.67	D0	53.67	92	70.67	CA	103.00	91
22.33	AE	37.00	30	54.00	1A	71.00	D6	104.00	9B
22.67	BE	37.33	32	54.33	12	71.33	DA	105.00	8B
23.00	BA	37.67	3A	54.67	00	71.67	DB	106.00	83
23.33	B8	38.00	31	55.00	08	72.00	D7	107.00	93
23.67	BC	38.33	33	55.33	06	72.33	D9	108.00	18
24.00	AC	38.67	3B	55.67	0E	72.67	D5	109.00	10
24.33	A8	39.00	3F	56.00	0F	73.00	D8	110.00	04
24.67	94	39.33	37	56.33	07	73.33	D4	111.00	0C
25.00	84	39.67	3E	56.67	17	73.67	20	112.00	16
25.33	8C	40.00	36	57.00	1F	74.00	28	113.00	1E
25.67	9C	40.33	34	57.33	0D	74.33	38	114.00	14
26.00	9E	40.67	4A	57.67	05	74.67	22	115.00	1C
26.25	8E	41.00	4B	58.00	1D	75.00	2A	116.00	F9
26.50	86	41.33	4E	58.33	15	75.33	39	117.00	FA
26.75	96	41.67	4F	58.67	FB	75.67	29	118.00	FD
27.00	0A	42.00	5F	59.00	FF	76.00	21	119.00	E9
27.25	02	42.33	5E	59.33	EB	76.33	23	120.00	FE
27.50	0B	42.67	5C	59.67	EF	76.67	2B	121.00	E8
27.75	03	43.00	5D	60.00	ED	77.00	27	122.00	FC
28.00	1B	43.33	54	60.33	EA	77.33	2F	123.00	43
28.25	13	43.67	55	60.67	EE	77.67	25	124.00	F2
28.50	09	44.00	50	61.00	EC	78.00	2D	125.00	F6
28.75	01	44.33	51	61.33	E6	78.33	3D	126.00	F8
29.00	19	44.67	AA	61.67	E2	78.67	35	127.00	5B
29.25	11	45.00	A6	62.00	E4	79.00	3C	128.00	5A
29.50	F3	45.33	A2	62.33	E0	79.33	2E	129.00	63
29.75	F7	45.67	B6	62.67	F4	79.67	2C	130.00	62
30.00	E7	46.00	B2	63.00	F0	80.00	26	131.00	77
30.25	E3	46.33	BB	63.33	60	81.00	24	132.00	76
30.50	E5	46.67	B0	63.67	64	82.00	49	133.00	52
30.75	E1	47.00	B9	64.00	74	83.00	48	134.00	53
31.00	F1	47.33	B4	64.33	70	84.00	4C	135.00	66
31.25	F5	47.67	BD	64.67	73	85.00	4D	136.00	67
31.50	61	48.00	A4	65.00	72	86.00	44	137.00	CC
31.75	65	48.33	A0	65.33	6C	87.00	45	138.00	CD
32.00	75	48.67	A9	65.67	7C	88.00	40	139.00	AB
32.25	71	49.00	AD	66.00	68	89.00	41	140.00	CF
32.50	6D	49.33	95	66.33	78	90.00	A7	141.00	CE
32.75	7D	49.67	85	66.67	7A	91.00	A3	142.00	DE
33.00	69	50.00	9D	67.00	7E	92.00	B7	143.00	BF
33.25	79	50.33	8D	67.33	6A	93.00	B3	144.00	DF
33.50	7B	50.67	89	67.67	6E	94.00	B1	145.00	DD
33.75	7F	51.00	99	68.00	6F	95.00	B5	146.00	DC
34.00	6B	51.33	88	68.33	C4	96.00	A5	147.00	AF
34.33	C1								

Notation:

delay (samples)	hex index
20.00	42
20.33	46
20.67	47
•	•
•	•
•	•

When a particular delay is chosen in the analyzer, the hexadecimal index associated with the delay, specified by this table, is transmitted to the synthesizer in the binary bitstream. For example to transmit a pitch delay of 20.00 the hexcode given by the table is 42. For this delay, the bitstream is transmitted in the following manner:

0100 | 0010

PD(n)-7 • • • PD(n)-0

Table 9. Pitch Delay Bit Stream Assignment

<0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	-1
0	0	-1	0	-1	0	1	0	0	0	0	0	0	0	-1	0	-1	-1	0	0
-1	0	-1	0	-1	0	-1	0	0	0	0	0	0	0	-1	0	0	0	1	0>
0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	-1	0
1	0	1	0	-1	1	0	0	0	0	0	-1	0	0	1	0	0	0	-1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1
0	1	-1	0	0	-1	-1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	-1	0	-1	-1	0	0	-1	0	1	0	0	0	1	0	0	0	0
0	-1	0	0	0	-1	0	0	0	0	0	0	-1	0	1	0	0	0	0	0
0	0	-1	1	0	0	0	-1	1	0	0	1	0	0	-1	0	0	0	0	0
0	-1	0	0	1	-1	0	0	1	0	0	0	0	0	0	1	1	1	0	0
1	1	0	0	0	-1	0	0	0	-1	0	0	-1	1	0	0	0	0	0	-1
-1	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	-1	-1	0	0	1	0
0	0	0	0	0	-1	0	0	-1	1	0	0	0	0	0	0	-1	0	0	-1
0	0	0	0	0	-1	0	0	-1	0	0	0	0	0	-1	-1	-1	0	-1	-1
0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	-1	0	0	0
0	0	0	1	-1	0	0	-1	0	-1	0	0	-1	0	0	1	0	0	0	0
-1	-1	0	-1	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0
-1	0	1	0	0	-1	0	0	0	-1	0	0	0	-1	0	0	0	0	-1	0
0	-1	-1	0	0	0	1	0	-1	0	0	0	0	0	1	0	0	0	0	0
0	0	1	1	0	0	0	1	1	0	0	0	-1	0	0	0	0	0	-1	1
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	-1	0	1	1	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	-1
-1	1	0	0	0	1	0	1	0	1	0	0	-1	0	0	0	0	0	0	0
0	-1	0	-1	0	0	0	-1	0	0	-1	0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	-1	-1	0	0	0	0	1	0	0	0	0	0	0
-1	0	-1	-1	0	0	-1	0	-1	0	0	0	0	0	1	0	1	0	0	0
0	0	0	1	-1	0	1	0	0	0	0	-1	0	0	0	0	0	0	1	0
0	-1	-1	0	0	0	0	0	0	1	0	0	0	0	1	-1	0	1	0	1
0	1	0	0	0	1	0	0	0	0	-1	0	0	0	-1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
0	0	0	0	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	1	0	0	0	0	0	0	-1	1	0	1	0	0	-1
0	0	0	0	0	0	0	0	0	1	-1	0	0	0	-1	0	0	0	0	-1
0	0	0	0	1	-1	0	-1	1	0	0	0	0	-1	0	0	0	1	0	0
0	0	0	-1	0	0	0	0	0	0	-1	0	-1	1	1	-1	0	0	0	1
0	1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	1	0
0	0	0	0	-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	-1	0	0	0	1	0	0	0	0	-1	0	1	-1	0	1	0	0	0
0	0	-1	0	0	0	1	0	-1	1	-1	0	0	1	0	0	0	0	-1	0
0	0	-1	0	0	0	0	-1	0	0	0	0	1	0	0	0	-1	0	0	0
0	0	0	1	0	0	0	0	0	0	1	-1	0	0	1	0	0	0	0	-1
0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1	0	0	1	0	0
0	0	1	0	0	1	0	0	-1	0	0	0	1	0	0	0	0	0	0	0
1	0	0	-1	0	0	-1	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	1	0	0	-1	1	-1	0	1	0	0	0
0	0	0	0	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	-1	0	0	0	0	-1	0
[0	0	(-1	0	-1	0	0	0	0	0	0	-1	0	0	-1	0	0	0	0	0
0	0	0	1	-1	0	0	0	0	-1	0	1	0	0	1	0	0	0	0	-1
0	0	0	0	1	0	0	0	0	-1	0	0	-1	0	0	0	0	0	0	1]
0	0)																		

(codeword 0) [codeword 1] <codeword 511>

Table 11. Stochastic Code Book

Bit	Description	Bit	Description	Bit	Description	Bit	Description
1	PG(4)-4*	37	PD(2)-0	73	PD(1)-4	109	CI(1)-2
2	PD(3)-4	38	CI(4)-1	74	CG(3)-2	110	PG(2)-1
3	LSP 1-1	39	LSP 9-0	75	LSP 7-1	111	CI(3)-7
4	CG(2)-4	40	CI(3)-8	76	CI(2)-7	112	LSP 4-0
5	CI(3)-3	41	PG(1)-4*	77	CI(3)-0	113	CI(2)-5
6	CI(1)-8	42	CG(2)-2	78	PD(2)-5	114	PD(1)-7*
7	PD(4)-0	43	PD(1)-3	79	LSP 4-1	115	PG(1)-0
8	LSP 8-0	44	LSP 6-1	80	CG(1)-0	116	CG(4)-4
9	PG(2)-3	45	CI(3)-4	81	PG(4)-3	117	LSP 5-0
10	CG(3)-0	46	CI(2)-2	82	LSP 9-1	118	PD(4)-2
11	PD(1)-5*	47	CG(1)-4	83	PD(3)-6*	119	CI(1)-3
12	LSP 3-3	48	PD(2)-3	84	CI(1)-4	120	CI(3)-1
13	CI(2)-3	49	LSP 1-2	85	CG(2)-1	121	LSP 7-2
14	CI(4)-4	50	PG(3)-2	86	LSP 6-2	122	CI(4)-2
15	PD(2)-1	51	HP-1*	87	CI(4)-3	123	PD(1)-1
16	LSP 10-0	52	PD(3)-1	88	PG(2)-2	124	PG(2)-4*
17	PG(1)-3	53	CG(4)-3	89	PD(4)-3	125	CG(3)-3
18	CG(4)-0	54	LSP 8-1	90	LSP 1-0	126	LSP 3-1
19	LSP 5-2	55	PG(3)-0	91	CG(4)-2	127	CI(1)-7
20	PD(3)-0	56	CI(2)-8	92	LSP 8-2	128	PD(3)-2
21	HP-0*	57	PD(4)-1	93	CI(2)-4	129	CI(2)-6
22	CI(1)-1	58	CI(4)-0	94	HP-2*	130	LSP 9-2
23	CI(4)-8	59	LSP 3-2	95	PD(2)-2	131	PG(4)-1
24	LSP 2-2	60	PG(2)-0	96	LSP 3-0	132	CG(1)-1
25	PG(3)-1	61	PD(1)-6*	97	PG(1)-2	133	PD(2)-4
26	PD(4)-5	62	CG(2)-0	98	CG(3)-4	134	HP-3*
27	CG(1)-3	63	CI(3)-6	99	LSP 10-2	135	LSP 6-0
28	CI(3)-5	64	LSP 10-1	100	CI(4)-5	136	PG(3)-3
29	LSP 7-0	65	PG(1)-1	101	CI(2)-0	137	CI(4)-6
30	CI(2)-1	66	CI(4)-7	102	PD(1)-2	138	PD(1)-0
31	PD(3)-7*	67	PD(3)-3	103	LSP 5-1	139	LSP 2-3
32	CI(1)-0	68	CG(1)-2	104	SP-0*	140	CG(4)-1
33	PG(4)-0	69	LSP 5-3	105	PG(4)-2	141	CI(3)-2
34	LSP 4-3	70	CI(1)-6	106	CG(2)-3	142	LSP 4-2
35	CG(3)-1	71	LSP 2-0	107	LSP 2-1	143	PD(3)-5*
36	CI(1)-5	72	PG(3)-4*	108	PD(4)-4	144	SY-0

Notes

i = 0 (Least Significant Bit) to Most Significant Bit   n = subframe number
LSP j-i = Line Spectral Parameter where j = 1 to 10 PD(n)-i = Pitch Delay
PG(n)-i = Pitch GainCI(n)-i = Code Book Index
CG(n)-i = Code Book GainHP-i = Hamming Parity
SP-i = Spare
SY-i = Synchronization
Order of Transmission is from bit 1 to bit 144* = Forward Error Corrected Bit

Table 12. Bit Assignment

Parameter	Bits Protected
Pitch Delay (1)	5, 6, 7
Pitch Delay (3)	5, 6, 7
Pitch Gain (1)	4
Pitch Gain (2)	4
Pitch Gain (3)	4
Pitch Gain (4)	4
Spare Bit	0

Table 13. Protected Data Bits

Parity Word	Invert Bit	Parity Word	Invert Bit
0	none	8	none
1	none	9	5
2	none	10	6
3	1	11	7
4	none	12	8
5	2	13	9
6	3	14	10
7	4	15	11

Table 14. Error Correction Decoding

III. REFERENCES

[1] Kemp, D., R. Sueda and T. Tremain, "An Evaluation of 4800 bps Voice Coders," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1989, p. 200-203.

[2] *Federal Standard 1016, Telecommunications: Analog to Digital Conversion of Radio Voice by 4,800 bit/second Code Excited Linear Prediction (CELP)*, National Communications System, Office of Technology and Standards, Washington, DC 20305-2010, 14 February 1991.

[3] Campbell, J., V. Welch and T. Tremain, "An Expandable Error-Protected 4800 bps CELP Coder," *Proceedings of ICASSP*, 1989, p. 735-738 (and *Proceedings of Speech Tech*, 1989, p. 338-341).

[4] Rahikka, D., T. Tremain, V. Welch and J. Campbell, "CELP Coding for Land Mobile Radio Applications," *Proceedings of ICASSP*, 1990, p. 465-468.

[5] Lin, D., "New Approaches to Stochastic Coding of Speech Sources at Very Low Bit Rates," *Signal Processing III: Theories and Applications (Proceedings of EUSIPCO-86)*, 1986, p. 445-448.

[6] Kleijn, W., D. Krasinski and R. Ketchum, "An Efficient Stochastically Excited Linear Predictive Coding Algorithm for High Quality Low Bit Rate Transmission of Speech," *Speech Communication*, October 1988, p. 305-316.

[7] Kroon, P. and B. Atal, "On Improving the Performance of Pitch Predictors in Speech Coding Systems," In *Advances in Speech Coding* (B. Atal, V. Cuperman and A. Gersho, Eds.). Kluwer Academic Publishers, Boston, 1991, p.321-328.

[8] Lin, D., "Speech Coding Using Efficient Pseudo-Stochastic Block Codes," *Proceedings of ICASSP*, 1987, p. 1354-1357.

[9] Chen, J.-H. and A. Gersho, "Real-Time Vector APC Speech Coding at 4800 bps with Adaptive Postfiltering," *Proceedings of ICASSP*, 1987, p. 2185-2188.

[10] Taniguchi, T., Y. Tanaka and R. Gray, "Speech Coding with Dynamic Bit Allocation (Multimode Coding)" In *Advances in Speech Coding* (B. Atal, V. Cuperman and A. Gersho, Eds.). Kluwer Academic Publishers, Boston, 1991, p. 157-166.

[11] Cox, R, W. Kleijn and P. Kroon, "Robust CELP Coders for Noisy Background and Noisy Channels," *Proceedings of ICASSP*, 1989, p. 739-742.

[12] Shoham, Y., "Constrained-Stochastic Excitation Coding of Speech at 4.8 kb/s," In *Advances in Speech Coding* (B. Atal, V. Cuperman and A. Gersho, Eds.). Kluwer Academic Publishers, Boston, 1991, p. 339-348.

[13] Campbell, J., V. Welch and T. Tremain, "The New 4800 bps Voice Coding Standard," *Proceedings of Military and Government Speech Tech*, 1989, p. 64-70.

[14] Trancoso, I. and B. Atal, "Efficient Search Procedures for Selecting the Optimum Innovation in Stochastic Coders," *IEEE Transaction on Acoustics, Speech, and Signal Processing*, March 1990, p. 385-396.

[15] Tremain, T., J. Campbell and V. Welch, "A 4.8 kbps Code Excited Linear Predictive Coder," *Proceedings of the Mobile Satellite Conference*, 1988, p. 491-496.

[16] Campbell, J., T. Tremain and V. Welch, "The DoD 4.8 kbps Standard (Proposed Federal Standard 1016)," *Advances in Speech Coding*, Kluwer Academic Publishers, Boston, 1991, p.121-133.